INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

Fast Wavelet Transforms and Seismic Compression

BY

FAOUZI M. KHENE

A Dissertation Presented to the
DEANSHIP OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY
In

ELECTRICAL ENGINEERING

May 2001

UMI Number: 3025132

UMI Microform 3025132

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

KING FAHD UNIVERSITY OF PETROLEUM AND MINERALS
DHAHRAN 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This dissertation, written by ***Mohammed Faouzi Khene*** under the direction of his Dissertation Advisor and approved by his Dissertation Committee, has been presented to and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements of the degree of **DOCTOR OF PHILOSOPHY IN *ELECTRICAL ENGINEERING*.**

Dissertation Committee

Dr. *Samir H. Abdul-Jauwad*
Dissertation Advisor

A. Siddiqi
Prof. *Abul Hassan Siddiqi*
Member

G. Korvin
Dr. *Gabor Korvin*
Member

Mahmoud Dawoud
Prof. *Mahmoud Dawoud*
Member

Dr. *Umar A. Al-Suwailem*
Member

Department Chairman

Dean of Graduate Studies

15/7/2001
Date

Acknowledgment

First and foremost, all praise is to Almighty *ALLAH* Who gave me the courage and patience to carry out this work, and peace and blessings be upon the *Prophet Muhamad.*

I am indebted to *King Fahd University of Petroleum and Minerals* for the facilities and support during this work. Appreciation is due to the *Electrical Engineering Department* for its support and to the Library facilities.

My deep appreciation goes to my supervisor Dr. *S.H.Abdul-Jauwad* for his excellent guidance and immeasurable support and encouragement at every stage of this work.

Thanks are also due to Prof. *M. Dawoud*, Dr. *O. Al-Suwailem*, Prof. *A.H.Sidiqqi* and Dr. *K. Gabor* for accepting to be in the thesis committee and for their interest, cooperation, advice and constructive criticism. Thanks are also due to Dr. *D. Mougenot* (CCG/ARGAS) for his valuable comments and support.

Acknowledgment is due to *Saudi ARAMCO* for providing the seismic data.

Last but not least, I wish to thank my parents for the unconditional support they have provided throughout my life, I could not have asked for more.

Contents

List of Figures

List of Abbreviations

ACF	autocorrelation function
AVO	amplitude versus offset
CCF	crosscorrelation function
CDF	Cohen-Daubechies-Feauveau
CMP	common midpoint
CSG	common source gather
CWT	continuous wavelet transform
DCT	discrete cosine transform
DFT	discrete Fourier transform
DMO	dip moveout
DPCM	differential pulse code modulation
DWT	discrete wavelet transform
EAGE	European association of geoscientists and engineers
ECSQ	entropy-constrained scalar quantizer
FFT	fast Fourier transform
FIR	finite impulse response
F-K	frequency-wavenumber
FWT	fast wavelet transform
HVS	human visual system
IBM	industrial business machine
IIR	infinite impulse response

IT	information theory
KLT	Karhunen-Loève transform
Lap	Laplacien pdf
LCT	local cosine transform
LPC	linear predictive transform
LVL	low velocity layer
MDL	minimum description length
MRA	multiresolution analysis
NMO	normal moveout
QMF	quadrature mirror filters
R(D)	rate-distortion function
RLC	run-length coding
SBC	subband coding
SEG	society of exploration geophysicists
SVD	singular value decomposition
TUSQ	threshold uniform scalar quantizer
TWT	two-way traveltime
WHT	Walsh-Hadamard transform
VAD	variable area display
VDD	variable density display
VM	vanishing moment
WPT	wavelet packet transform
WST	wavelet series transform

DISSERTATION ABSTRACT

FULL NAME: FAOUZI M. KHENE
TITLE OF STUDY: Fast Wavelet Transforms and Seismic Compression
MAJOR FIELD: Electrical Engineering
DATE OF DEGREE: May 2001

Fast and efficient wavelet-based seismic compression strategies are proposed. The compression scheme consists of three units, namely a fast 2-D wavelet transform, an adaptive quantizer and an entropy encoder. Both orthogonal and biorthogonal wavelet systems have been used. The fast wavelet transform is implemented using both the lattice structure and the lifting scheme. However, the lifting scheme is more versatile and well suited for large seismic datasets. Indeed, it reduces the computation complexity by half, it saves the auxiliary memory usage thanks to its time-domain in-place implementation and it is immediately invertible with exactly the same complexity as for the forward transform. The wavelet coefficients are then quantized using a statistical adaptive threshold uniform scalar quantizer (TUSQ). The zero-bin size is determined for each detail subband using two efficient and powerful statistical techniques based on the *SURE* principle and the *Birgé-Massarat* strategy. The design of the TUSQ outside the dead-zone is performed using the *minimum description length* (MDL) principle. The lossless binary encoding stage consists of a *run-length coding* (RLC) followed by a *Huffman* coder. It is similar to the one prevailing in the JPEG2000 standard. Finally, the proposed compression strategies are tested with real seismic data from the *Red Sea* provided by *Saudi ARAMCO*.

ملخص بحث

درجة الدكتوراه في الفلسفة

الاسم :محمد فوزي خان
عنوان الرسالة :تحويلات الموجة الصغيرة السريعة و الضغط الزلزالي
التخصص :الهندسة الكهربائية
تاريخ التخرج :مايو ٢٠٠١

يقترح هذا البحث استراتيجيات سريعة و فعالة مبنية على تحويلة الموجة الصغيرة لضغط البيانات الزلزالية. تتكون خطة الضغط من ثلاث وحدات و هي على التوالي: تحولية الموجة الصغيرة السريعة ذات البعدين ،*كوانتيزر* تكيفي و آلة تشفير *انتروبيا*. وقد استعملت نظم الموجات الصغيرة المتعامدة والمتعامدة المزدوجة، كما نفذت تحولية الموجة الصغيرة السريعة باستعمال هيكل *الإطار الشبكي* و *الخطة الرافعة*. و تعتبر الخطة الرافعة أكثر فعالية لتعدد جوانبها و تناسبها مع البيانات الزلزالية ذات الحجم الكبير. بالفعل، يقلل استعمال الخطة الرافعة تعقيد الحساب إلى النصف و يوفر استعمال الذاكرة الإضافية بفضل خطة حسابية زمنية تمكن من استبدال البيانات الأصلية بالبيانات المحصلة كما تنفد التحويلة العكسية بنفس السهولة. يكمن دور الوحدة الثانية من خطة الضغط في التقسيم الكمي لمعاملات التحويلة باستعمال *كونتايزر* إحصائي عددي ذو عتبة (**TUSQ**). و يتم حساب طول العتبة للمعاملات المنتمية إلى نفس الشريط باستخدام تقنيتين إحصائيتين فعالتين، الأولى مبنية على مبدأ **SURE** و الثانية مبنية على إستراتيجية **Birgé-Massarat** . أما تصميم *كونتايزر* **TUSQ** خارج المنطقة الميتة فيتم بفضل استعمال طريقة *الحد الأدنى لوصف الطول* (**MDL**) . أما الوحدة الأخيرة من خطة الضغط فتتكون من التشفير الثنائي الدقيق المبني على مشفر *هفمان*. وأخيرا، طبقت إستراتيجيات الضغط المقترحة على بيانات زلزالية فعليه جمعت في البحر الأحمر ووزعت من قبل أرامكو السعودية.

Chapter 1

INTRODUCTION

Seismic compression technology plays an important role in managing seismic data in a world of ever increasing size of datasets. Seismic data compression allows for a greater flexibility in managing local or remote server disk space and reducing network traffic [89],[90]. It also encourages innovative interpretation workflow to leverage the vast information content in 3-D seismic data while maintaining or exceeding current productivity levels. Ten years ago, 3-D surveys were acquired primarily over proven fields to optimize development programs [74]. Nowadays, 3-D mega-surveys over vast areas are routinely evaluated [11],[66],[97]. Elsewhere, service companies have dramatically increased acquisition capacity as acquisition technologies make it cost effective to collect large-scale 3-D surveys covering thousands of square kilometers marine acquisition [98].

Furthermore, as repeated 3-D (4-D) seismic is utilized more and more to monitor reservoir performance [48],[105], and as the seismic industry embraces multi-component data [99] to better image the subsurface or characterize the reservoir, geophysicists will undoubtedly require new technologies to manage and interpret even greater volumes of data. Faster computing platforms, improved communication networks, and data storage technology have contributed to accommodate the seismic data explosion, yet explorationists continue to face interpretation workflow challenges and *information technology* (IT) issues.

The present dissertation aims at developing strategies to quickly and efficiently compress seismic data. The adopted compression strategy consists of three units, namely a fast 2-D wavelet transform, an adaptive quantizer and an entropy encoder.

1.1 Compression Strategy

Compression may be divided into *lossless* and *lossy* techniques. Lossless compression yields an exact reconstruction (*noiseless*) of the data. It is used either for compressing text files, e.g. seismic headers or symbols for binary coding. Lossy compression achieves much higher compression rates than lossless one [102] by allowing some *distortion* or error. As seismic data has already an amount of inherent errors due to the floating point format and subsequent data processing, lossy techniques are more suitable for seismic compression [35], [103].

All compression[1] techniques consist of applying a *signal expansion* or *transformation* to the data prior to quantization [16],[41],[49],[62]. The rationale of transformations is to decorrelate the data and to compact its energy. For 1-D data like speech, correlation exists only along one direction, in both the time and the frequency domains. Waveform coding techniques, which exploit the correlation in the time-domain, like *differential pulse*

---

[1] In the sequel, the term compression means lossy compression.

code modulation (DPCM) [50] and in the frequency domain, like *subband coding* (SBC) [76] are commonly used. For high-dimensional data, the decorrelation may be achieved by using either a *separable* or a *non-separable* transform [68],[76],[104].

Irrespective of the data dimension, correlation encompasses *redundancy* and *irrelevancy* [76]. Lossless coding relies on the redundancy and exploits unequal symbol probability distributions and thus predictability. A trivial example of redundancy is a binary string of all zeros or ones. It carries no information and can be coded with essentially one bit for its length. On the other hand, lossy coding deals with irrelevancy. Data irrelevancy occurs for instance in the visualization of grayscale images with high dynamic range. It is an experimental fact that for monochrome images, 6 to 8 bits of dynamic range is the limit of *human visual system* (HVS) [69] as any extra bits do not add perceptual value and may be eliminated [63]. The great variety of compression algorithms differ mainly in their approaches to extracting and exploiting redundancy and irrelevancy.

1.1.1 Fast wavelet transform

In case one knows precisely the correlation present in a dataset, e.g. the *correlation matrix*, the *Karhunen-Loève transform* (KLT) constitutes the optimal transform [50]. The basis vectors are the *eigenvectors* of the correlation matrix. This yields the best information packing results because the eigenvectors diagonalize the latter. However, this theoretical optimal representation, has several practical disadvantages [50],[104]:

- In most cases the correlation matrix is not known
- The algorithm to calculate the eigenvectors of a matrix has a cubic complexity, i.e.

requires $\mathcal{O}(N^3)$ operations in general

- The transform is a quadratic algorithm, i.e. $\mathcal{O}(N^2)$
- The basis functions are data dependent

To overcome the above shortcoming, one possible candidate is the *fast Fourier transform* (FFT). Its real version, namely the *discrete cosine transform* (DCT) has been until recently the prevailing decomposition engine of commercial coders, e.g. JPEG[2] [62]. However, the exponential basis functions that are perfectly localized in frequency, are totally delocalized in time.

As most signals have both local frequency and spatial correlation, the *discrete wavelet transform* (DWT) is more appropriate. Indeed,

- It takes into account the time-frequency characteristics of the data
- It is independent of the dataset
- It is capable of removing the correlation for large and general datasets
- The transform is linear[3], i.e. $\mathcal{O}(N)$

In the present dissertation, efficient factorization techniques based on the *lifting scheme* are proposed in order to further reduce the computation complexity by half, i.e. $\mathcal{O}\left(\frac{N}{2}\right)$ and to save the auxiliary memory.

---

[2] The new JPEG2000 standard , incoporates however, the DWT [52].

[3] The FFT is logarithmic, i.e. $Nlog(N)$.

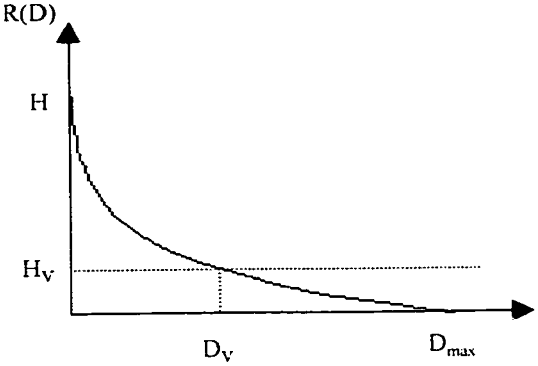


Figure 1-1: A typical rate-distortion function of a discrete source. The threshold D_V is the lowest observable distortion corresponding to the visual entropy H_V.

1.1.2 Adaptive quantizer

In the present work, the joint design of the quantizer and the subsequent entropy encoding unit is adopted. This joint design results in the so-called *entropy-constrained scalar quantizer (ECSQ)*. It yields a *variable length quantizer*, in which the different quantization choices are assigned variable *codelengths*. The performance of the quantizer in conjunction with the entropy encoder characterizes the operational *rate-distortion* R(D) function of the source [4]. The theoretical R(D) function characterizes the fundamental lossy compression limit theoretically attainable, and it is rarely known in analytical form except for a few special cases, such as for a *Gaussian* source. A typical rate-distortion function for a discrete source[4] is depicted in figure 1-1. The data irrelevancy threshold is given by the *visual entropy* [76]. From figure 1-1, it can be inferred that the distortion D decrease exponentially as the rate R increases. Under high-resolution conditions, it can be shown

---

[4] For an analog source the zero distortion entropy is infinite.

that the optimal ECSQ is a uniform one, whose average distortion is close to the theoretical bound attainable also known as the *Shannon* bound [50],[104]. For low bit rate coding, a uniform quantizer with a *dead-zone* (or *zero-bin*) in the central quantization bin is often used [76]. This means that the all-important central bin is wider than the other bins. This turns out to be more efficient than having all bins of the same size. The performance of dead-zone quantizers is nearly optimal for memoryless sources even at low rates. However, the dead-zone is in general selected as twice the uniform step size, i.e. 2Δ [76], which lacks flexibility especially if the source consists of the empirical wavelet coefficients. In order to enhance the quantizer flexibility, an efficient and adaptive dead-zone uniform scalar quantizer based on the *wavelet shrinkage* [29] and *minimum description length* (MDL) techniques [79] is adopted.

1.1.3 Entropy encoder

Quantization results in an ensemble of integers, with a lot of them contiguously repeated. To get an efficient binary encoding, these numbers are converted to *symbols* using *run-length coding* (RLC). The basic idea behind RLC is to map a sequence of numbers into a sequence of symbol pairs (*run*;*value*). For example the sequence (1110010000) can be mapped to these symbols $\{(3;1),(2;0),(1;1),(4;0)\}$. Different implementations might use a slightly different format as in the case of the thresholded detail wavelet coefficients, which consist of long runs of zeros. The symbol *value* is used to code only the value of the nonzero quantized coefficients and the symbol *run* is used to code the number of zeros preceding each non-zero number [68],[104].

The last step of the adopted compression scheme consists of the *entropy encoding unit*, which aims at achieving the entropy rate of the quantizer. The two popular entropy encoding methods are the *Huffamn coding* and the *arithmetic coding*. The former associates efficiency with simplicity, which gives it the merit to be incorporated in the JPEG2000 standard [52]. It will also used in the present dissertation.

1.2 Seismic Imaging

Oil and gas are usually buried deep within the earth, often kilometers below the surface. Most of the easy or shallow oil has been found. A number of exploration methods are available [85]. The most common are the *seismic reflection*, the *gravity*, the *magnetic*, and the *electrical* techniques. Sometimes *spectral reflectivity* and *temperature measurements* are also made.

Seismic reflection[5][85],[109], consists of generating acoustic waves within the earth interior and measuring reflected waves at different subsurface interfaces. In this sense, seismic reflection is a very sophisticated version of the echo sounding used in *radar* systems or in *bathymetric* surveys. In addition to examining the times of arrival of these signals, seismic reflection processing extracts also information about the subsurface from the (relative) *amplitude* and *phase* of the reflections. The complex subsurface structures are rather interpreted in terms of *reflectors*, i.e. boundaries separating material with differing elastic properties than in terms of layers.

A typical seismic exploration scenario is illustrated in figure 1-2. For land surveys,

---

[5] As opposed to the *seismic refraction*, which is not dealt with in this dissertation.

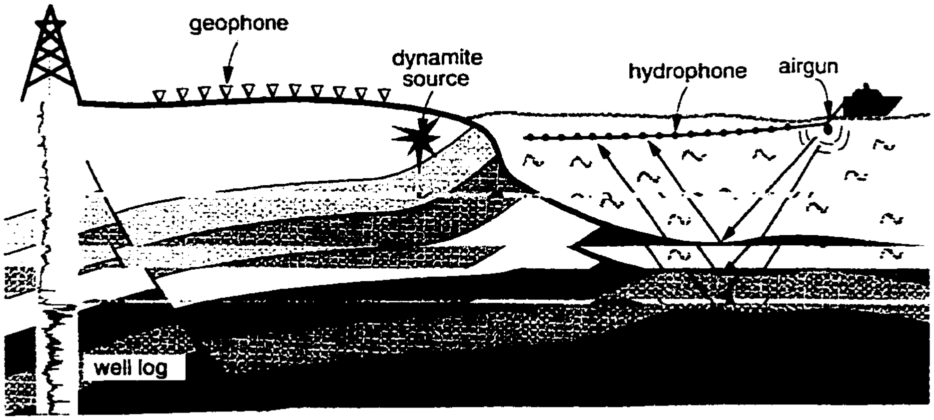


Figure 1-2: Illustration of the seismic reflection method (Courtesy [26]).

a single seismic energy source, e.g. *dynamite* or *vibrators*, generates a wavefield that is recorded by numerous array of *geophones* laid along lines (2-D seismic) or over large areas (3-D seismic). For each shot a number of signals, e.g. 24, 48 or 96, are multiplexed and recorded digitally on a magnetic tape. A geophone converts particle velocity to voltages and produces a (seismic) *trace*, i.e. a time series of usually 2 and 15[6] seconds long. Reflected signals contain frequencies from a few hertz to a few hundred hertz, and field data are usually sampled every 2 or 4 *ms*. For purposes of suppression of certain types of undesirable energy, i.e. random and surface noise, a group of geophones (may be as many as 100!)[7] are used at each receiver location and the output is summed to form the signal from one *receiver station*, which is transmitted in a multiplexed mode with signals from others stations to a *recording truck* (land acquisition) or *seismic vessel*

---

[6] for some special crustal surveys.

[7] *Schlumberger* has recently developed a new recording system, the *Q-system*, in view of replacing the geophones group with one single geophone. Composition of the single geophone records by digital array forming provides a noise free, high resolution seismic trace.

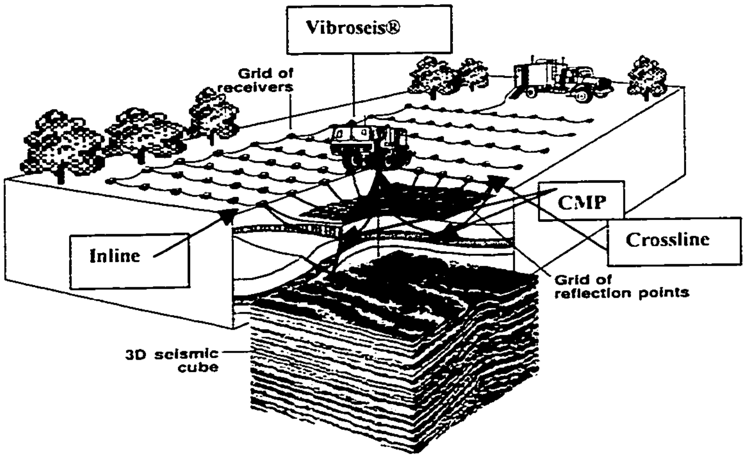


Figure 1-3: 3-D seismic measures the reflectivity of the subsurface within a volume whose surface corresponds to the array of receivers. Buried structures are reconstituted precisely from the continuous tracking of reflections throughout the seismic volume. (Courtesy [74])

(marine acquisition). In *marine surveys*, an *air-gun* is used to generate the wavefield and the receivers, called *hydrophones*, are towed to a vessel in an *off-end* spread geometry. In addition to these remote and global prospecting techniques, the earth is also probed directly and locally using *logging techniques* as shown in figure 1-2.

Seismic data acquired simultaneously from a swath of receivers laid out over several square kilometers is called a 3-D seismic data set (see figure 1-3). After processing, the latter provides a *seismic cube* like the one represented by figure 1-4. A seismic cube is defined with three geometric coordinates, namely the shot direction, the CMP's and time. As illustrated in figure 1-4, cross sections can thus obtained along the CMP, the line or the TWT coordinates yielding an in-line, a crossline or a time slice section respectively. Moreover, a fourth dimension, say the evolution time, may be added to make seismic data a production tool by comparing repeated surveys (4-D seismic) [105].

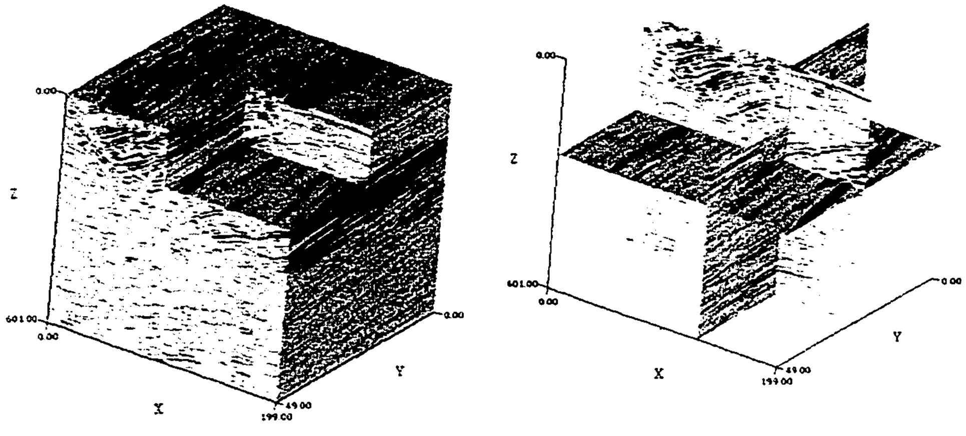


Figure 1-4: Seismic cube and cross-sections (Stratton (Texas) gas field dataset [44]).

1.2.1 Seismic record

The entire collection of traces associated with one source is called a *common shot gather (CSG)*. The latter can be either *off-end spread* (as in figure 1-2) with all the receivers lying to one side of the source, or *split-spread* with the source being at the center of the receivers spread.

The CSG is commonly defined in terms of the source coordinate s, the receiver coordinate g, the midpoint coordinate m, and the half-offset coordinate h as described by the basic diagram of figure 1-5. Seismic data acquisition is done in shot-receiver coordinates (s, g) and after the initial signal processing, the data are in general transformed to midpoint-offset coordinates (m, h). The required coordinates transformation is achieved by sorting the data into midpoints between the source and the receiver locations. Those traces with the same midpoint location are grouped together making up a *common midpoint (CMP)* gather.

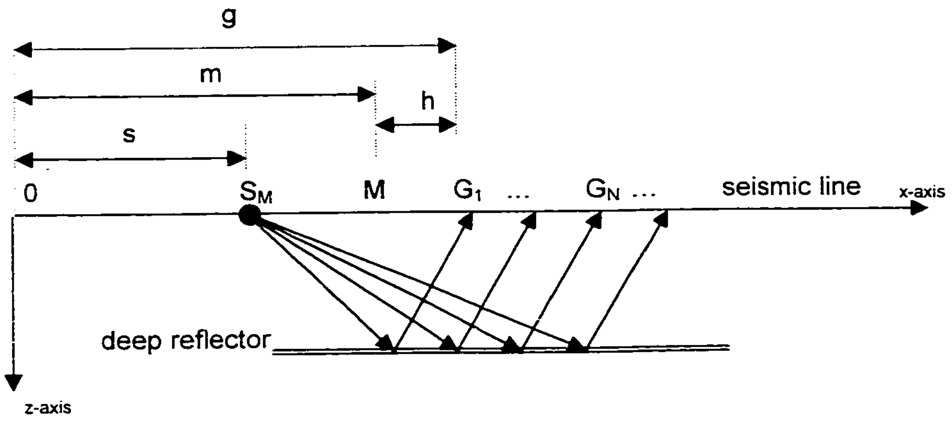


Figure 1-5: Definition of seismic processing coordinates: source coordinate s, geophone coordinate g, midpoint coordinate $m = \frac{g+s}{2}$, and half-offset coordinate $h = \frac{g-s}{2}$.

In a seismic record, seismic waves do not appear as such but rather induce characteristic *patterns* in the CSG, whose shapes are mainly a function of the distance of the receiver from the *source* (i.e. the offset) and the velocity of propagation of the waves. These patterns are:

- *Reflections*
- *Multiples*
- *Critical refractions*
- *Ground roll and air-blast*
- *Direct arrivals*
- *Diffractions*

Some of the above patterns are schematically depicted in figure 1-6. Moreover, all real

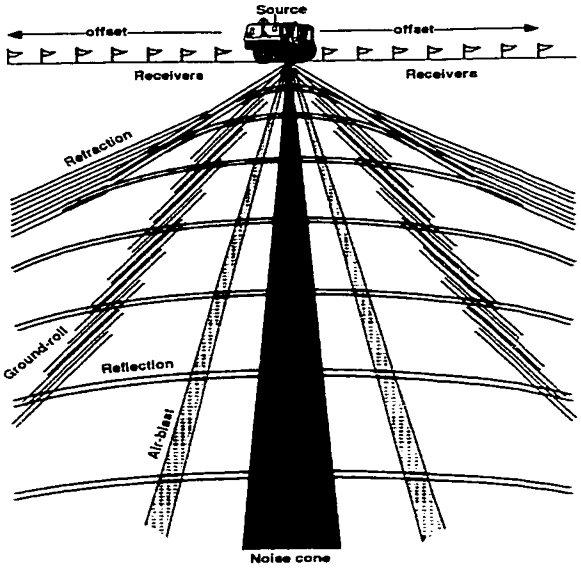


Figure 1-6: Common seismic signatures on a split-spread shot point.

data contain many wiggles that seem random in size, shape, and location, with no obvious pattern from trace to trace. The sources of these wiggles could be:

- *Wind noise, electronic glitches, pumping stations, vehicles etc.*
- *Weak overlapping primary and multiple reflections*
- *Organized scattered noise*

Whatever their source, almost all of above effects are undesirable noise as far as standard processing is concerned. A noisy CSG acquired using the split-spread configuration appears in figure 1-7.

1.2.2 Seismic data display

There are three common modes for displaying seismic records:

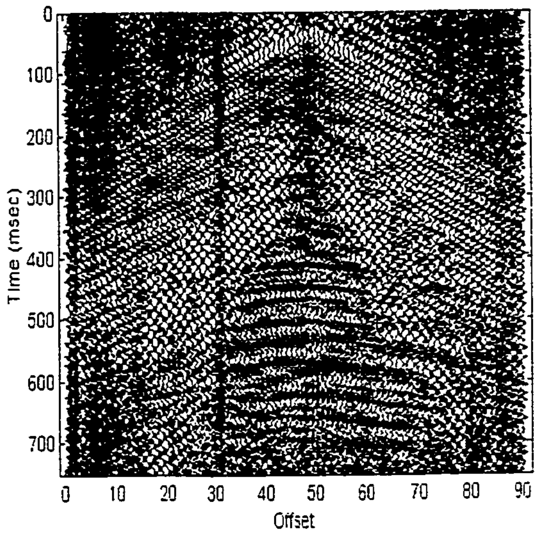


Figure 1-7: Correlated vibroseis CSG showing a series of reflections, ground-roll and random noise [109].

Variable area display (VAD) - The amplitude of the traces of a seismic record are shown as an horizontal deflection from baseline. In general, the *wiggle plot* of figure 1-8 comes with area fill on positive peaks to yield the VAD of figure 1-9. This kind of display is useful for close-up views where it is important to inspect the waveforms.

Variable density display (VDD) (Figure 1-10) - The amplitude of the traces of a seismic record are assigned a gray value between black (max) and white (min). Usually 256 shades of gray (8-bits amplitude) are used. It is often the best way to preview large quantities of data.

Color display - The amplitude of the traces of a seismic record are assigned an RGB color code or one of the many other coloring methods.

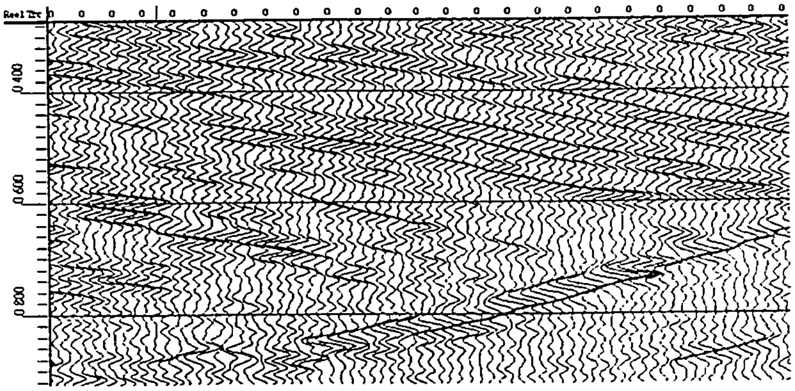


Figure 1-8: Wiggle plot display mode (Data courtesy *Saudi ARAMCO*).

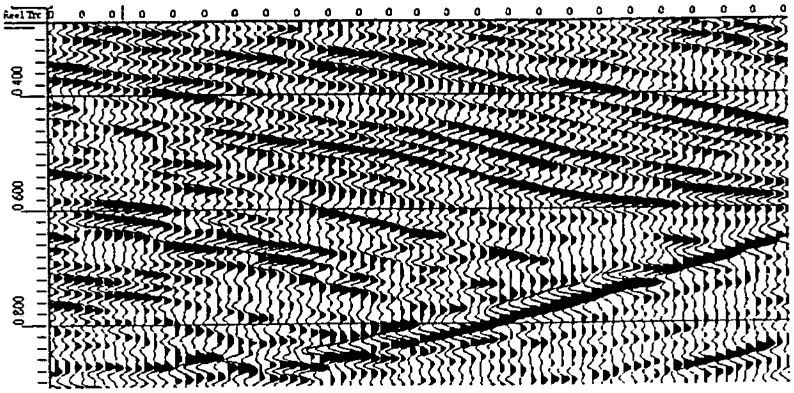


Figure 1-9: Variable area display (VAD) mode.

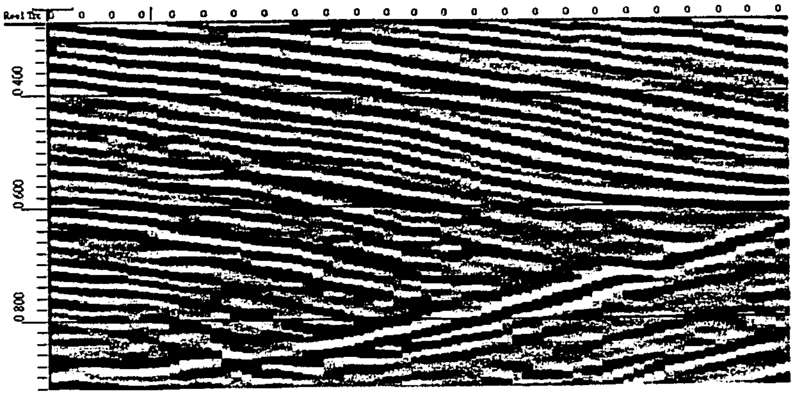


Figure 1-10: Variable density display (VDD) mode.

1.2.3 Seismic data format

Modern instrumentation and data acquisition systems record fairly accurate facsimiles of the ground response to incoming reflected energy [75]. The recording instruments filter, amplify, multiplex, digitize, and write on a magnetic tape the signals issued from receivers groups. Digital seismic recording systems have a dynamic range $> 100dB$. Reflection amplitudes decay about $100dB$ in the first 4 seconds of recording, owing primarily to attenuation losses along the travel path. Consequently, the amplifier gain level is continuously changing during the recording phase to preserve traces amplitudes [75].

Currently seismic field tapes are recorded in a number of standard formats but the most used one is the *Society of Exploration Geophisicists* (SEG)-Y format [3] that is based on the IBM character set and numeric format (the prevailing at that time). There are other formats to record seismic data such as the SEG A, SEG B, SEG C, which are data sequential and the SEG D, which is trace sequential. The SEGY format is a demultiplexed format composed of three parts: the *reel header*, the *trace identification headers* and the *seismic data*. In the tape reel header, one can distinguish between two parts: a 3200-byte EBCDIC skeleton arranged as 40 card images, and a 400-byte binary coded block, of which 60 are assigned and 340 left available for future expansion. The latter is carrying information about the reel and the line numbers, the sampling period, the number of traces, and the number of the *CMP* fold, to cite a few. The first 240-bytes of each trace record are binary encoded trace identification information. Mainly, these are the trace sequence number within a line, the original field record number, the trace number within original field record, the number of samples in the trace and the sampling period.

The remaining information in the record pertains to the sample values, almost invariably in IBM floating point format, which employs a 7-bit exponent and 24-bit mantissa.

1.2.4 Seismic data processing

Seismic signal processing usually begins with the *demultiplexing* of the CSG gathers resulting in seismic traces in sequential order followed by *trace editing* to remove noisy traces with transient glitches or mono frequency signals that are due to the transmission lines.

There are three central stages in seismic data processing, each of which aims at improving seismic imaging by bringing back the reflections to their true vertical and lateral positions with respect to the subsurface reflectivity, i.e. velocity and density contrasts. In their usual order of application, these are *deconvolution*, *stacking* and *migration*.

Deconvolution - Due to the low-pass filtering effect of the earth, the source signal that started as a spike (white spectrum) is reduced to a broader pulse that is of lower energy and with a net phase shift. *Spiking deconvolution* or *whitening* aims at improving the temporal resolution by compressing the effective source wavelet contained in the seismic trace to a spike. Deconvolution is usually performed after *true amplitude recovery* (TAR) corrections, which compensate for the *spherical divergence* and data editing.

Stacking - It consists essentially in summing traces of variable offset belonging to the same midpoint. Prior to that, each trace on a CMP is (dynamically) corrected for *normal moveout* (NMO). Under the assumption that all the reflectors are flat and level[8] and for a given reflector, NMO aims at reducing the *two-way traveltime* (TWT) for the slant path

---

[8] In the case of dipping reflectors, a further *dip moveout* (DMO) correction is required.

to that of the zero-offset (i.e. vertical) path. After the NMO correction, the event due to a horizontal reflector will line up on all the traces of the CMP gather at the same zero-offset TWT. The traces can be added together, and such reflections are thus reinforced, while the noise (including multiples) is attenuated. This process is called *stacking* and the resulting output trace is called the *stacked trace*. One stacked trace results from each CMP gather and the collection of all the stacked traces makes up the *stacked section*.

Migration - The stacked section does not provide information about the position of the different reflectors as the reflections are assumed to come from the CMP. However, in most of the prospecting areas the reflectors could not be either horizontal or with a homogeneous velocity. *Time migration* techniques help then in increasing the lateral resolution moving dipping events to their true subsurface location and by collapsing diffractions. In this respect, migration is considered as a *spatial* deconvolution that improves the spatial resolution of seismic images.

1.3 Seismic Compression

1.3.1 State-of-the-art

The pioneering work in lossy seismic compression is the one of *Wood* [108]. *Wood* has developed a sequency-domain technique that, when applied to test data, enabled signals to be recorded in less than $1/28^{th}$ of their original size. His approach was to truncate the *Walsh transform* of a signal (sequency band limit the data), and then to perform memoryless variable-length coding based upon some nominal probability distribution. However, the impressive results obtained with this technique were at the expense of significant distortion in the data because of the importance of both weak events (the reason for maintaining high dynamic range in recording) and the timing of events in interpreting a seismic record. Therefore, although the reconstructed signals were satisfactory for remote plotting, which was the main concern of *Wood*, their limited quality would degrade the performance of more sophisticated seismic analysis. In addition, the storage reduction achieved by this scheme could not be determined without actually performing the encoding. Consequently, implementing this approach effectively when the available storage is constrained would be difficult: either the data would have to be encoded several times to determine an appropriate cutoff for the sequency band limiting, or a low cutoff and thus a high distortion level would have to be accepted to ensure sufficient compression.

Later, *Stingall* and *Panagos* [86] have investigated two easily time-domain schemes for compression during acquisition. The first involved applying memoryless encoding after differencing consecutive samples of the data. This reduced the storage required by a factor of four. However, integrated white noise was introduced in reconstructing the signals, resulting in sequences which became progressively less accurate with time. Consequently, later portions of the data were poorly represented. The high-pass filtering proposed as a cure, would eliminate much significant low-frequency information from the data. The second scheme investigated in the study involved filtering the signals with fixed-length polynomial predictors with data-independent coefficients. Compression was accomplished by only storing those signal samples, which could not be successfully predicted from a fixed number of previous signal values. Prediction was deemed successfully when an *a priori* error bound was satisfied. These approaches all share two serious shortcomings. First, the performance of any of these schemes with respect to a norm is not known. Thus, it is difficult to determine whether these methods trade distortion for compression efficiency or even whether they are capable of working well on data ensembles other than those actually tested. Second, it is not clear that these techniques perform well when only moderate or small distortion is allowed.

The first application of coding techniques used in speech processing and data communication to seismic compression, was performed by *Bordley* for archiving marine seismic data [7]. *Bordley* applied a *linear predictive coding* (LPC) in performing a high-fidelity seismic data compression. The basic idea of the approach is that if a signal is sufficiently predictable, i.e. the energy in the whitened signal is sufficiently small compared to the original signal energy, then the dynamic range of the whitened signal is significantly less than the dynamic range of the original data. Therefore, fewer bits are required to represent the whitened signal and the filter parameters than to represent the waveform directly. Then a quantizer is necessary to reduce the precision of the filtered data to the desired wordlength. *Bordley* investigated the degradation introduced in compressing an ensemble of ocean bottom seismic refraction signals for a reduction in the wordlength up to 2 bits and calculated absolute limits on the minimum distortion possible. These estimates of the distortion bounds could be used for real-time monitoring of the performance of the encoder. However, the presented technique is performed on a trace-by-trace basis and does not take into consideration the coherency of the data in the space direction.

The first formal application of transform-based coding of seismic data was conducted by *Spanias et al.* [87]. They investigated the performance, for storage purpose, of transform coding of seismic data recorded by the *Norwegian regional seismic array* (NRSA). The NRSA was installed to monitor underground nuclear explosion in order to verify treaties on nuclear testing. The coding capacity of four transform has been examined, namely the *Karhunen-Loève transform* (KLT), *discrete Fourier transform* (DFT), *discrete cosine transform* (DCT) and *Walsh-Hadamard transform* (WHT). A subband approach was adopted in all the transform domains and a uniform quantization rule was applied to encode the transform coefficients. However, the seismic data used in quite different from the seismic data that stem form seismic prospecting because the recorded seismic signals are essentially one-dimensional P-waves.

The first noticeable interest by the seismic industry in wavelet transform-based compression is marked by two main talks at the 62^{nd} *annual international meeting of the society of exploration geophysicists* (SEG). *Steinlein* and *Hjelle* [89] discussed a method for the compression of seismic sections based on *splines*. B-spline-based compression has been previously used by *Helle* and *Daehlen* [45] for structural maps compression. The authors accomplished data compression by keeping only the coefficients from the components that give a significant contribution to the reconstruction. By significant, the authors meant that the distance between the original seismic section and the reconstruction is less than a predefined tolerance in some convenient norm. Experiments show that satisfactory interpretation quality is readily obtained with storage requirements in the range of 1 to 2 bits per sample. *Luo* and *Schuster* [67] as their part, applied the *wavelet packet transform* (WPT) to compress earthquakes and borehole data. They achieved compression by discarding the small coefficients of the transform. Their approach, even though theoretically superior to the wavelet transform, is very expensive to implement because the prohibitive number of bases candidates.

At the 63rd SEG meeting, *Bosman* and *Reiter* [8] have formally introduced the *discrete wavelet transform* (DWT) in seismic data compression. Their main concern was the quantification of the errors introduced by the compression scheme. They concluded that smoothing processes such as *migration* or *stacking* would mitigate the errors while local or derivative processes such as *deconvolution* would increase any errors present. In the same context, another study has been conducted by *Reiter* and *Heller* [77] using NMO-corrected CMP gather rather than CSG. The authors pointed that the spatial redundancy present in seismic data is best exploited for compression by 2- and 3-dimensional wavelet transforms. Elsewhere, they showed that the compression noise is not correlated along the horizontal stacking trajectories. This result may explain why the pre-stack data can tolerate higher compression ratios than the stacked data.

The trade-off between the compression ratios using higher dimensions and random access of the seismic traces in the compressed dataset was a concern of *Chen*'s Msc. thesis [15]. *Chen* dealt with seismic compression for both storage and transmission purposes. He also applied *local cosine transform* (LCT) and DWT to both 1- and 2-dimensional data. He did not, however, investigate the effect of the choice of wavelet bases and the quantization rules on the overall performance of the compression scheme.

At the 57th *European association of geoscientists and engineers (EAGE)* meeting, *Hewlett* and *Hatton* [46] discussed the aspect of seismic data compression for CD-ROM storage, whereas *Hall et al.* [43] emphasized the impact of compression noise on the interpretability of the final product. The benefit of using additional dimensions on seismic data compression has been further discussed by *Donoho et al.* [32]. The authors proposed a high dimensional wavelet transform compression of a seismic volume prior to satellite transmission [90]. They used a separable 3-D wavelet transform and assumed a separable *Markov model* for the data. More importantly, they expect further improvements in data-compression quality if the quantization rules are matched to the transform stage.

The crystallization of the different seismic compression aspects started in 1996 at the 66^{th} SEG Meeting. *Reiter* [78] performed a comparative study to investigate the effect of 1-D, 2-D and 3-D wavelet compression on a single dataset and evaluated the compression errors in terms of the $L^1$and L^2 metrics. He concluded that, provided that there will be a correlation between these metrics and *data fidelity*, the compression noise may decrease substantially, with additional dimension at higher rates. *Dean* and *Mosher* [23] discussed the effect of compression on interpretation projects. They concluded that compression should yield improved results in a timely manner by having the capability to access simultaneously many type of data such as stacked, migrated and *amplitude versus offset* (AVO) attribute sections. *Brudely et al.* [10] evaluated the effect of compression on images obtained from a *Kirchoff* depth migration. They considered a seismic cube of 2000(time) x 200(space)x 6(shots) resulting from a collection of common source experiments conducted on the EAGE/SEG salt model. They used a hybrid 3-D transform, which consists of a separable 2-D DWT applied to the time-space plane and a KLT applied along the shot direction. According to the authors, the use of a 3-D separable DWT was prohibited because the small shot number dimensionality. Long *biorthogonal wavelet filters* were used followed by a *scalar quantization* and *Huffman* coding. The authors concluded that high compression ratios require data with smaller amplitude dynamic range. They suggested the eliminating of noise bursts and the scaling of the amplitudes. to account for wavefront spreading, prior to compression. *Diller et al.* [27] highlighted another important aspect of seismic compression, which consists of the need of decompression standard routines to enable exchange of compressed datasets. Also, *Mosher* [72] invited the seismic community to work on a standard for data compression. *Vermeer et al.* [103] proposed to use compression employing a *local cosine transform* (LCT). Their method has been used in a marine survey to transmit shot gathers from the vessel to the processing center via a mere 16 *KB/sec* line. The received data were used for velocity analysis, selection for a processing modules and parameters. *Guo* and *Burrus* [42] raised the point of phase preservation in seismic compression. They proposed a shift-invariant transform they called the *self-adjusting wavelet transform*. They also proposed a variant of the embedded *zerotree wavelet compression technique*, which is well adapted for progressive transmission. However, unlike natural image coding, progressive decompression and transmission is inadequate for seismic data because only full precision data is used for inspection, procession and interpretation. In [33], *Donoho et al.* investigated the important aspect of seismic data compression optimization. The authors proposed to derive a set of quantization rules experimentally by analyzing a number of typical datasets. Their approach is limited, however, because it is not clear that it will work well with any dataset, elsewhere, the authors did not investigated the optimization of either the transform or the encoding stages.

The optimization aspect of seismic compression has been investigated further in 1997. We mention the works of *Rosten et al.* [83], *Waldemar et al.* [106], and *Vassiliou* and *Wickerhauser* [101]. *Rosten et al.* emphasized the importance of proper filter design and frequency partitioning in subband coding of seismic stack sections previously dealt with in [82]. They applied an $8x8$ separable parallel nonunitary 32-tap FIR filter bank that has been optimized to maximize the coding gain under the perfect reconstruction constraint. The optimization, however, assumes that a separable auto-regressive process of order 1 and 2 along the space and time directions respectively, models the seismic section. The detail subbands have been quantized using the same *threshold uniform scalar quantizer* with a zero bin equals twice the bin size and encoded using an *arithmetic coder*. Because of the strong correlation within the lowest resolution subband, the latter has been coded separately using two different approaches. The first approach consists of using a third order closed-loop *differential pulse code modulation* (DPCM) quantization. The second one, consists of using a 3-stage separable octave-band tree nonunitary perfect reconstruction FIR filter bank, followed by a uniform threshold quantizer and a class-wise dynamic arithmetic coder allocation. The authors concluded that their coder is suitable for storage and transmission applications. *Waldemar et al.* proposed the *singular value decomposition* (SVD) as an alternative to the DWT. The authors have not mentioned however the implementation issues of the SVD decomposition, which requires the calculation of the eigenvectors of the input data. Finally, *Vassiliou* and *Wickerhauser* conducted an experimental study to select the optimal combination of the wavelet filters and the quantization methods to obtain the best quality for a given compression ratio. They tested three different biorthogonal wavelet filters for four different levels of decomposition and three different quantization schemes on two different datasets. The authors concluded that, although their study is not exhaustive, the best combination seems to be the $(9,7)$-*biorthogonal filter*[9] decomposition to levels $(6,4)$ in time and space directions coupled with uniform quantization.

The tendency of the 69th SEG and the 61st EAGE meetings was the development of metrics for seismic compression evaluation. This appears in the contributions of *Donoho et al.* [34] and [35], which are related to earlier works of *Ergas et al.* discussed at the EAGE 58th [36] and 66th SEG [37],[38] meetings. As their part, *Donoho et al* stressed on the necessity of developing seismic data compression quality measures and proposed a diagnostic procedure to establish the acceptability of seismic compression for prestack datasets [34]. Elsewhere, *Donoho et al.* mentioned the formation in 1997, of an industry-wide consortium designated the *seismic compression diagnostic initiative (SCDI)* whose main objective is the development of diagnostic tools that will allow users of compression to determine whether a desired compression ratio is safe to use with a given dataset. The SCDI has begun carrying out a range of tests on a number of datasets but no definitive results have yet been made public.

1.3.2 Seismic compression guidelines

- The success with which seismic data can be compressed depends on the dataset being compressed and the purpose that the data serves. Furthermore, the effect of

---

[9] Also known as the FBI wavelet filter.

pre-processing on the compression ratios can be considerable: more compression can be achieved by attenuating random noise and by discarding strong events that are not required in the decompressed data. However, if for instance shot records are transmitted for quality control the scope for pre-processing is limited: the quality control operator will want to inspect all constituents of the data. On the other hand if data are transmitted to be processed a refraction mute, *F-K filtering*, etc., may be desirable. Although data compression offers many benefits to interpretation productivity and IT resource management, interpreters must balance these benefits with the inherent loss in data fidelity during the compression process. Since compression is lossy, interpreters must first consider the degree of fidelity suitable or required to achieve interpretation goals [34],[35].

- 1-D compression techniques exploit only the correlation of data along the time axis as they are performed trace-by-trace by using dynamic window and appropriate basis function of the transform. Selecting an appropriate dynamic window, which divides the traces into various size segments, depends on the data features. Note that seismic data is time-variant, i.e. there are significant changes with time in both amplitude levels and bandwidth in spite of the gain corrections applied to data throughout the processing stages. However, seismic data is not stochastic because the earth reflectors do not change randomly and so the arrival times. Techniques using stochastic assumptions may break down at high ratios.

- The amount of 1-D compression achieved would be less than the one attained with a 2-D compression. However, the access to a given trace in the compressed domain requires the decompression of that trace only. In order to benefit from the trace-to-trace correlation, also called *coherency*, a 2-D compression scheme is used. At the time the computation means were not very developed, some flexibility in trace accessing was gained by dividing the data into strips.

- Nowadays, it is common to compress whole seismic sections using 2-D compression techniques. As most of these have been developed for image compression, their applications to seismic data should be done with some provision. Firstly, the underlying components of a seismic image, are waveforms whose attributes, e.g. amplitude and phase, should be preserved as accurately as possible. Secondly, a seismic section has a high degree of *anisotropy*, large amplitude *imbalances* and discontinuities. Thirdly, decompressed seismic images are subject to post-interpretation computation, which is not the case of images used for multimedia applications.

- Data volume size can be reduced simply by throwing out traces or entire lines. Moreover, data are often divided into small sub-volume data sets for interactive viewing and manipulation. Although decimation strategies can lead to as much as 4 : 1 disk storage savings, this benefit comes at the expense of losing dynamic range, or of loosing regional context while viewing small subsets of data. Reducing the data volume with an appropriate degree of compression can maintain fault clarity, whereas conventional decimation strategies will smear the seismic data across fault zones.

1.4 Scope of the Thesis and its Novelties

As discussed in the historical background, a number of methods for compression seismic data that span the spectrum from simple to sophisticated have been applied. Seismic compression has undergone dramatic developments but it is still facing new challenges especially with the advent of 3-D and 4-D seismic. In the present contribution, efficient compression techniques are proposed in view of solving the problems faced by the seismic industry in managing ever increasing data sets. The global approach adopted is based on joint design of the compression units instead of optimizing the transformation or the quantizer independently. The rest of the dissertation is organized as follows:

In *chapter 2*, the *time-scale analysis* is introduced and the wavelet transform and its variants are defined. First the traditional *Fourier* transform and its time-varying version, namely the *short-time Fourier transform* (STFT) or windowed *Fourier* transform are reviewed. This will clear a path to the wavelet analysis that shares many properties with the *time-frequency analysis*, notably the atomic decomposition concept and the *Heisenberg's* uncertainty principle. The *discrete wavelet transform* (DWT), which is used for data compression, is first defined through the *continuous wavelet transform* (CWT). Indeed, the CWT helps to understand wavelets and to shed the light on their appealing properties such as the *time-frequency localization* and the *frequency selectivity*. The compression mechanisms of the DWT are best described in terms of the *wavelet series transform* (WST), which is conceptually similar to the *Fourier* series expansion, but is more adaptive and diverse. Then, the theory underlying the DWT, namely the *multiresolution analysis* (MRA), is presented. The role of the MRA is threefold. Firstly, it explains the hierarchical multiscale decomposition of real life signals. Secondly, it leads to the implementation of the wavelet multiresolution decomposition/reconstruction using the *fast wavelet transform* (FWT). Thirdly, the MRA offers the ideal framework for the design of versatile wavelet families endowed with very desirable properties such as *symmetry*, *regularity*, and *compactness of support*.

This is the concern of *chapter 3*, which is contended with the theory and the design of wavelet bases. Wavelet systems can be either *orthogonal* or *non-orthogonal*. An orthogonal MRA system is one in which the scaling functions are orthogonal to one another; the wavelets are orthogonal to one another, and each of the wavelets is orthogonal to every coarser scaling function. The most famous orthogonal wavelets are the *compactly supported Daubechies* wavelets [20]. Orthogonality leads to a robust and symmetric implementation of the DWT. It places however strong limitations on the possibilities of the transformation stage. In order to gain flexibility, orthogonality needs to be sacrified either partially or totally. This yields the *semi-orthogonal* and the *biorthogonal* wavelet systems respectively. In the semi-orthogonal MRA, the wavelets at a given resolution are at least orthogonal to all coarser scaling functions, though not necessarily orthogonal to each other. Unfortunately, the semiorthogonal systems suffer from some limitations when extended to higher dimensions, which is not the case for the biorthogonal systems. The rationale of the biorthogonal systems is to work with two dual multiresolutions rather than only one. The primal scaling functions are thus orthogonal to the dual wavelets and the primal wavelets are orthogonal to the dual scaling functions. The *Cohen-Daubechies-Feauveau* (CDF) class [17] constitutes the most famous biorthogonal wavelets. From the implementation point of view, the biorthogonal system can be factored using the *lifting scheme* into elementary cells yielding thus the most efficient implementation of the DWT [22]. Wavelet systems can be constructed by seeking a smooth time function which fulfills the MRA requirements [53]. However, this approach does not lend itself to a systematic design of large wavelet families. Elsewhere, it is only restricted to the semi-orthogonal and orthogonal instances. In addition, the constructed orthogonal systems do not have a compact support. On the contrary, the design of wavelet families using the framework of MRA is very efficient as it is based on digital filters design theory. This method has been recently initiated by *Daubechies* [17],[19], and leads to the design of large families of orthogonal and biorthogonal systems endowed with many appealing properties such as regularity, orthogonality, symmetry, and compactness of support. Finally, in order to apply the DWT to seismic images, the extension of the 1-D MRA to the 2-D case is considered.

Chapter 4 bridges the gaps between the theory and the implementation of the DWT. Indeed, computational complexity is what makes the difference between an impractical and a useful algorithm like the *fast wavelet transform* (FWT). The DWT is implemented efficiently by exploiting the scaling property of the MRA theory where the ensuing recursion constitutes the fundamental principle behind the calculation of the DWT. Moreover, it unravels the connections between the multiresolution analysis and filter banks [100]. The fundamental algorithm for the computation of the DWT, namely the *Mallat algorithm*, is first discussed before establishing the connections between the DWT and the *two-channel filter bank*. Then the different approaches to extend the 1-D DWT to the 2-D case are explained. In general, it is common to use a separable 2-D DWT where the filter bank is applied to each dimension individually. This can be achieved following two fashions. The first one consists of applying the 1-D DWT to each row of the input data matrix, e.g. pixels in the case of digital images. Next, the wavelet transformed rows are treated as if they were themselves a new image and the DWT is applied again but to each column. This leads to the so-called *standard wavelet decomposition*. The second type of separable 2-D DWT is called the *non-standard decomposition*. It alternates between operations on rows and columns, by first performing one step of pairwise 1-D DWT on the pixel values in each row of the image matrix. Next, a similar pairwise 1-D DWT is applied to each column of the result. To complete the transformation, the same process is repeated recursively only on the quadrant containing averages in both directions. As the non-standard wavelet expansion is computational efficient and simple [91], it is exclusively used in the present work. The last section of this chapter concerns borders corrections. In theory the DWT operator maps an infinite dimension signal $f(t)$ to an infinite set of wavelet coefficients, but seismic data do not extend indefinitely in time or space but are finite length signals. Consequently, the DWT operator may be truncated to a finite and invertible one or alternatively, the finite signal may be extended to an infinite one. In both cases, the resulting transform differs at the boundaries, while remaining the same for data far enough from the boundaries. Moreover, truncating the wavelet operator means designing the so-called *wavelets on the interval* [18], which have proven to be effective in theory but are not satisfactory from a practical point of view [25],[68]. It is often preferable to use simple schemes based on signal extension at the boundaries. These methods include: *zero-padding*, *periodic* and *symmetric extensions*.

Chapter 5 deals with the most efficient implementation of the DWT, namely the *lifting scheme* [94]. The lifting scheme has many advantages over the direct filter bank implementation. Firstly, it constitutes a very powerful technique to factorize every unitary and non-unitary FIR filter bank into a finite sequence of simple cells known as *ladder structures*, reducing thus the complexity by half, i.e. $\mathcal{O}(N/2)$ [22]. Then it allows for an in-place implementation of the FWT, which save any auxiliary memory usage. This turns out to be of the outmost importance when compressing large seismic data sets. Secondly, every lifted DWT is immediately invertible with exactly the same complexity as for the forward transform. Lastly, the lifting scheme exhibits an inherent parallelism that can be exploited when multiprocessors workstations are used to perform the on-line compression/decompression of seismic data. The concept of the lifting scheme in the general context is first explained, then its tight connection with paraunitary filter banks is highlighted using the polyphase representation. Finally, the factorization of *any* FIR filter bank was it orthogonal or biorthogonal into ladder structures is performed using the lifting scheme.

Chapter 6 tackles the challenging problem of wavelet coefficients encoding. Two statistical encoding strategies are proposed. They are based on two efficient and powerful concepts, namely the *wavelet shrinkage principle* developed by *Dohono* and co-workers [28],[29],[30], and the *minimum description length* (MDL) developed by *Rissanen* [79]. The proposed encoding engine is an adaptive version of the *threshold USQ* (TUSQ) [104], [107]. Indeed, instead of selecting the zero bin in an *ad hoc* manner as twice the bin size [76], an adaptive wavelet shrinkage approach is used to determine the best zero-bin based on the statistics of each detail subbands population. The design of TUSQ outside the threshold zone, i.e. the number of bins and their size, is based on new results from *Kolmogorov* theory [64]. The *Kolmogorov* theory is complex but its practical variant, i.e. the MDL is more tractable yet very efficient. The rationale of the MDL principle is to select a model with low complexity, i.e. having a simple representation or a reasonable number of parameters, and fitting well the quantized coefficients. Lastly, the lossless binary encoding stage consists of a *run-length coding* (RLC) followed by a *Huffman* coder. It is similar to the one prevailing in the JPEG2000 standard [52].

Finally, the proposed compression strategies are applied to a real seismic data set provided by *Saudi ARAMCO*. The data set consists of a time-migrated section of the *Midyan Basin* in the *Red Sea* (KSA) [73]. First, the problem of borders corrections is dealt with using both orthogonal and biorthogonal wavelet systems. Then, the ability of the orthogonality, the symmetry, the smoothness and the support size of the 2-D wavelet bases to compact the energy and to reduce the entropy of the input data is evaluated. Visual as well as statistical tools, such as the correlation functions and the histograms, are used. After experimentally selecting the most suitable wavelet bases for transforming the seismic data, the problem of wavelet shrinkage is tackled. As the design of the proposed adaptive quantizer is based on the statistics of the detail subband coefficients, the distribution of the coefficients is first analyzed using histogram fitting technique and quantile-to-quantile plots. Then an adaptive dead-zone for the TUSQ is designed for every detail subband using the SURE principle [30] and the *Birgé-Massarat* strategy [6] respectively. Finally, in view of demonstrating the ability of wavelet shrinkage to filter out seismic coherent noise, the challenging problem of ground-roll removal is considered using a different dataset described in the reference manual of *Yilmaz* [109] and available from the *Colorado School of Mines* anonymous ftp site at ftp.cwp.mines.edu.

Chapter 2

WAVELET TRANSFORMS

In this chapter, the *time-scale analysis* is introduced and the wavelet transform and its variants are defined. First, the traditional *Fourier* transform and its time-varying version, namely the windowed *Fourier* transform are reviewed. This will clear a path to the *wavelet analysis* that shares many properties with the *time-frequency analysis*, notably the atomic decomposition concept and the *Heisenberg* uncertainty principle. The *discrete wavelet transform* (DWT), which is the most suitable for data compression, is introduced through the *continuous wavelet transform* (CWT). Indeed, the CWT allows to define wavelets and to shed light on their appealing properties such as the *time-frequency localization* and the *frequency selectivity*. The compression mechanisms of the DWT are better explained in terms of the *wavelet series transform* (WST), which is conceptually similar the *Fourier* series expansion, but is much more adaptive and diverse. Then, the theory underlying the DWT, namely the *multiresolution analysis* (MRA) is presented. The MRA offers a powerful and systematic framework for the construction and the characterization of various wavelet bases [68], [104] that can be explicitly endowed with properties such as symmetry, regularity, and approximation power. Finally, the MRA allows the derivation of the *fast wavelet transform* (FWT), a fast and efficient algorithm for the computation of the DWT.

2.1 Wavelet Analysis

2.1.1 Time-frequency analysis

Let $f(t) \in L^2(\mathbb{R})$ be an analog signal with finite energy, i.e. $\|f\|^2 \equiv \int_{-\infty}^{+\infty} |f|^2\, dt < \infty$, then its *spectrum* is given by the *Fourier transform*,

$$\hat{f}(\omega) \equiv \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt \longleftrightarrow f(t) \equiv \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega) e^{i\omega t} d\omega \tag{2.1}$$

In practice, the *Fourier* transform is not suited for computing the spectral information of a signal because it requires all previous as well as future information about the signal to evaluate its spectral density at a single frequency ω. Elsewhere, since the *Fourier* transform of a time-varying signal is a function independent of time, it does not register frequencies varying with time. In order to achieve *time-localization* of the spectral characteristics of a time varying signal, a window function $\gamma(t)$ is introduced into the *Fourier* transform to form the so-called *short-time Fourier transform* (STFT) or *windowed Fourier transform*.

The STFT is a multidimensional functional that maps a one-dimensional signal to a two dimensional time-frequency plane [104]. Mathematically, the STFT is expressed as,

$$\hat{f}(\omega, \tau) \equiv \int_{-\infty}^{+\infty} [f(t)\gamma^{*}(t-\tau)]e^{-i\omega t} dt \tag{2.2}$$

where $\gamma^{*}(.)$ stands for the complex conjugate. Equation (2.2) means that multiplying the signal $f(t)$ by the window function $\gamma(t)$ before taking its *Fourier* transform has the effect of restricting the spectral information of the latter to the domain of influence of the window function. Furthermore, using translates of the window function $\gamma(t-\tau)$ along the time axis to cover the entire time domain, leads to the analysis of the spectral information in localized neighborhoods in time of the time-varying signal.

2.1.2 Heisenberg's uncertainty principle

In order to achieve a high degree of localization in both *time* and *frequency*, we need to select an arbitrary window function $\gamma(t) \in L^{2}(\mathbb{R})$ with sufficiently narrow time and frequency widths. Theoretically, $\gamma(t)$ could be chosen such that both $\gamma(t)$ and $\hat{\gamma}(\omega)$ have rapid decay, i.e. γ is well localized in time, while $\hat{\gamma}$ is well localized in frequency. This is not true in practice. Indeed, let the signal behavior within the STFT be illustrated using the concepts of signal *time duration* σ_t and *frequency bandwidth* σ_ω defined respectively as,

$$\sigma_t^2 = \frac{\int_{-\infty}^{+\infty} (t-\mu_t)^2 \left|f(t)\right|^2 dt}{\|f\|^2} \quad \text{with} \quad \mu_t = \frac{\int_{-\infty}^{+\infty} t \left|f(t)^2\right| dt}{\|f\|^2} \tag{2.3}$$

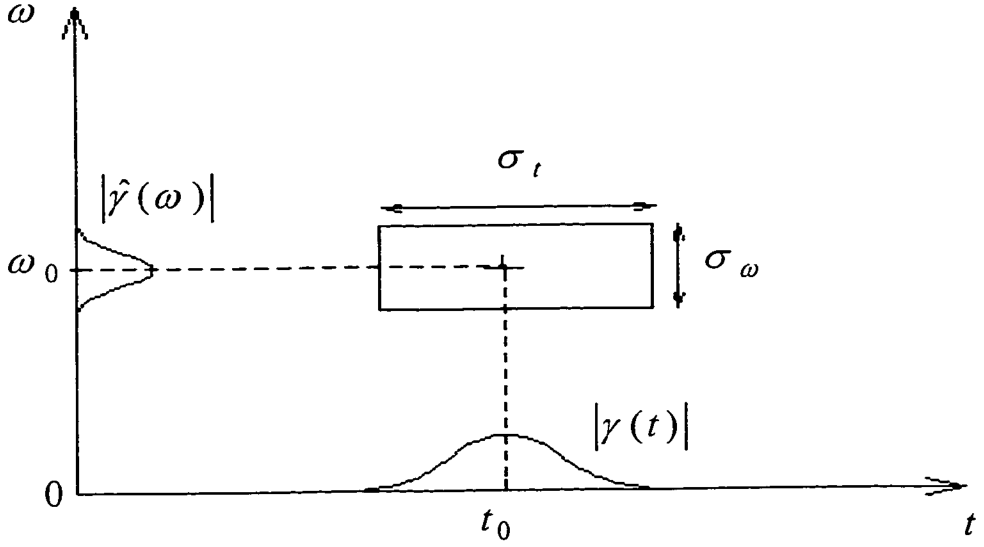


Figure 2-1: Time-frequency representation of an atom $\gamma(t)$.

$$\sigma_\omega^2 = \frac{\int_{-\infty}^{+\infty}(\omega-\mu_\omega)^2\left|\hat{f}(\omega)\right|^2 d\omega}{\left\|\hat{f}\right\|^2} \quad \text{with} \quad \mu_\omega = \frac{\int_{-\infty}^{+\infty}\omega\left|\hat{f}(\omega)\right|^2 d\omega}{\left\|\hat{f}\right\|^2} \tag{2.4}$$

Now, suppose that the window $\gamma(t)$ is centered at $t = 0$ and its *Fourier* transform $\hat{\gamma}(\omega)$ is centered at $\omega = 0$, then equation (2.2) gives a measure of the behavior of the signal $f(t)$ in the vicinity of $[t-\sigma_t, t+\sigma_t]$x$[\omega-\sigma_\omega, \omega+\sigma_\omega]$ as depicted by figure 2-1. Furthermore, by virtue of the *Parseval theorem*, the time-duration σ_t is tied to the frequency bandwidth σ_ω via the *Fourier* transform. Thus, they could not be both arbitrary small and their product must satisfy the *Heisenberg uncertainty principle*, i.e.

$$\sigma_t\sigma_\omega \succeq \frac{1}{2} \tag{2.5}$$

that imposes a theoretical lower bound on the area of the *time-frequency* window. The equality holds when $\gamma(t)$ is a *Gaussian* function, i.e.

$$\gamma(t) = \left(\frac{1}{\rho^2\pi}\right)^{\frac{1}{4}} e^{-\frac{t^2}{2\rho^2}} \leftrightarrow \hat{\gamma}(\omega) = \left(\frac{\rho^2}{\pi}\right)^{\frac{1}{4}} e^{-\frac{\rho^2\omega^2}{2}} \tag{2.6}$$

which yields the tightest time-frequency window with

$$\sigma_t = \frac{\rho}{\sqrt{2}} \quad \text{and} \quad \sigma_\omega = \frac{1}{\sqrt{2}\rho} \tag{2.7}$$

The STFT is then referred to as the *Gabor*[1] transform.

2.1.3 Time-scale analysis

With the STFT any features of the signal involving time intervals much shorter than the width of the window are underlocalized in time. Similarly, any features of the signal involving time intervals much longer than the window width are overlocalized in time. This makes the STFT an inefficient tool for analyzing regular time behavior that is either very rapid or very slow relative the window width.

The *wavelet*[2] *transform* solves both of these problems by using a *wavelet kernel* $\psi_{ab}(t)$, where *scaling* replaces *modulation* to achieve *frequency localization*. Conceptually, this means changing the size of the time window for different degrees of localization by reciprocally varying the size of the frequency window at the same time, so as to keep the area

---

[1] The discrete inverse STFT is also named in honour to Dennis *Gabor*, as the *Gabor* Expansion.

[2] Techniques such that wavelet shaping or wavelet deconvolution that are routinely used in seismic data processing should not be confused with the wavelet transform.

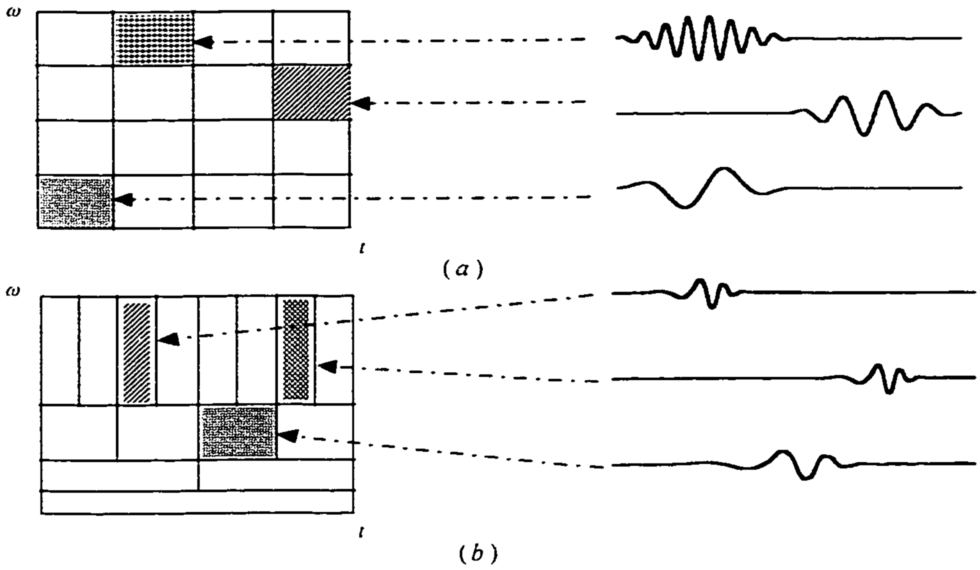


Figure 2-2: Constant versus proportional bandwidth time-frequency tilling: (a) STFT (b) CWT.

of the window constant. This implies a trade-off between the time- and the frequency-localization that is achieved by directly windowing the signal instead of its *Fourier* transform and by scaling the window function appropriately to change its time window width. Unlike the STFT atoms that yield a constant bandwidth tilling of the time-frequency plane (figure 2-2-(a)), the wavelet atoms yield a tiling in a *proportional-bandwidth (constant-Q)* fashion as shown in figure 2-2-(b).

A *wavelet system* is a basis-like system $\{\psi_{ab}(t)\}$ having a continuous set of labels. In $L^2(\mathbb{R})$, the double-indexed wavelet system $\psi_{ab}(t) \equiv |a|^{-\frac{1}{2}}\,\psi(\frac{t-b}{a})$ is constructed by continuously dilating and translating a prototype function $\psi(t)$, where a and b are the scale and the shift parameters respectively. As $\psi(t) \in L^2(\mathbb{R})$, the L^2-*norm* is also preserved. The prototype function $\psi(t)$ goes by many names in the wavelet literature like the *basic* *function*, the *mother wavelet*, the *wavelet function*, the *analyzing wavelet*, or the *wavelet* in short.

Wavelet analysis encompasses various transforms such as the *continuous wavelet transform* (CWT), the *wavelet series transform* (WST), the *discrete wavelet transform* (DWT) and the *wavelet packet transform* (WPT). As far as compression is concerned, only the DWT is of interest in the present dissertation. The CWT is extremely redundant yet a powerful analyzing tool but it is introduced here to get a better understanding and characterization the DWT. Indeed, dilations are not defined over discrete sequences and discrete wavelet bases have a more complicated structure [80]. The regularity of a discrete sequence is not well defined either, which makes it more difficult to interpret the amplitude of the wavelet coefficients. On the contrary, there exist several theorems relating the amplitude of the CWT to the local regularity of a signal [20],[68]. Therefore, the theory of continuous time functions gives asymptotic results for discrete sequences with sampling intervals decreasing to zero.

If the wavelet $\psi(t)$ verifies certain existence conditions and if the scale factor a is chosen sufficiently close to unity ($a \approx 1$) and b sufficiently small ($b \approx 0$), there exists a discrete wavelet frame, which is highly redundant but allows a stable inverse wavelet transform. A very interesting case occurs when the CWT samples are taken on a *dyadic grid*, i.e. the scale factor $a \equiv 2^j$ and the shift factor $b \equiv k2^j$, $\{j, k\} \in \mathbb{Z}$, yielding a critically sampled discrete frame. The resulting wavelet system $\{\psi_{jk}(t) \equiv 2^{-\frac{j}{2}}\psi(2^{-j}t - k)\}$ forms an *orthonormal basis* rather than a mere frame [20]. However as a is not close to unity and b is not small enough, the above orthonormal wavelet basis could not stem from the discretization of a generic continuous frame, but is rather generated through *multiresolution analysis* (MRA) [53]. The MRA offers a powerful and systematic framework for the construction and characterization of various wavelet bases [68], [104], that can be explicitly endowed with many interesting properties such as symmetry, regularity, and approximation power. Moreover, the MRA allows the derivation of the *fast wavelet transform* (FWT), a fast and efficient algorithm for the computation of the DWT.

2.2 Continuous Wavelet Transform

The CWT was first introduced by *Goupillaud*, *Grossman* and *Morlet* around 1984, but received full attention from the signal processing community when *Daubechies*, *Meyer* and *Mallat* established connections between the DWT and *filter banks* [70].

2.2.1 Forward transform

The CWT is a multidimensional mapping of a signal $f(t) \in L^2(\mathbb{R})$, depending on a continuous time (or space) parameter t (or x) to a 2-D function $\check{f}(a, b)$ depending on a scale (or dilation) parameter a and a shift (or temporal translation) parameter b. Mathematically, the CWT is defined as the integral transform on $L^2(\mathbb{R})$ [19] ,[53], [68],

$$\check{f}(a,b) \equiv W\{f,\psi\}(a,b) = \int_{-\infty}^{+\infty} f(t)\psi_{ab}^{*}dt \tag{2.8}$$

where the wavelet $\psi \in L^2(R)$. The wavelet family $\psi_{ab}(t)$ is defined as

$$\psi_{ab}(t) \equiv |a|^{-\frac{1}{2}} \psi \left(\frac{t-b}{a} \right) \tag{2.9}$$

with $a, b \in \mathbb{R}$ and $a \neq 0$. The factor $|a|^{-\frac{1}{2}}$ is introduced to ensure energy preservation such that if $\int_{-\infty}^{+\infty} |\psi(t)|^2 \, dt = 1$ then $\int_{-\infty}^{+\infty} |\psi_{ab}(t)|^2 \, dt = 1$. The width of the wavelets $\psi_{ab}(t)$ is controlled by the scale parameter a such that,

1. If $|a| < 1$ then the wavelets are compressed,
2. If $|a| > 1$ then the wavelets are dilated (or stretched)

Besides. the position of the stretched or compressed wavelets is controlled by the shift parameter b. These cases are illustrated by figure 2-3, using the *Ricker* wavelet[3].

Remark 1 *The CWT can be interpreted differently leading to two interesting point of views. It can thought of as the correlation of the signal $f(t)$ with the wavelet $\psi(t)$ and in this instance the wavelet coefficients are a measure of the degree of similitude between the two. Alternatively. the integral (2.8) can be considered as the inner product of the signal $f(t)$ with the wavelets $\psi_{ab}(t)$, i.e.*

$$\check{f}(a, b) \equiv \langle f, \psi_{ab} \rangle = f * \psi_{ab} \tag{2.10}$$

hinting at the connection between the CWT and filtering theory.

---

[3] also known as *Mexican hat* or *Sombrero*.

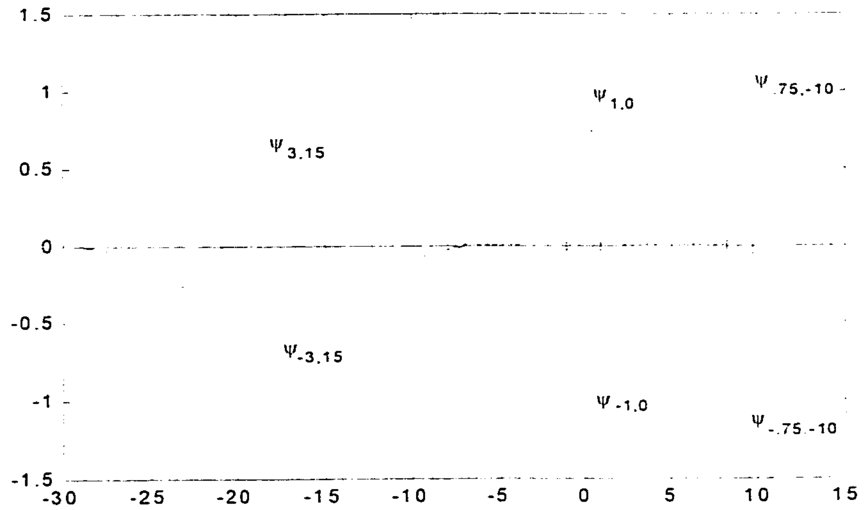


Figure 2-3: A family of wavelets formed by scaling and shifting a *Ricker* mother wavelet. Negative scaling results in reflected wavelets.

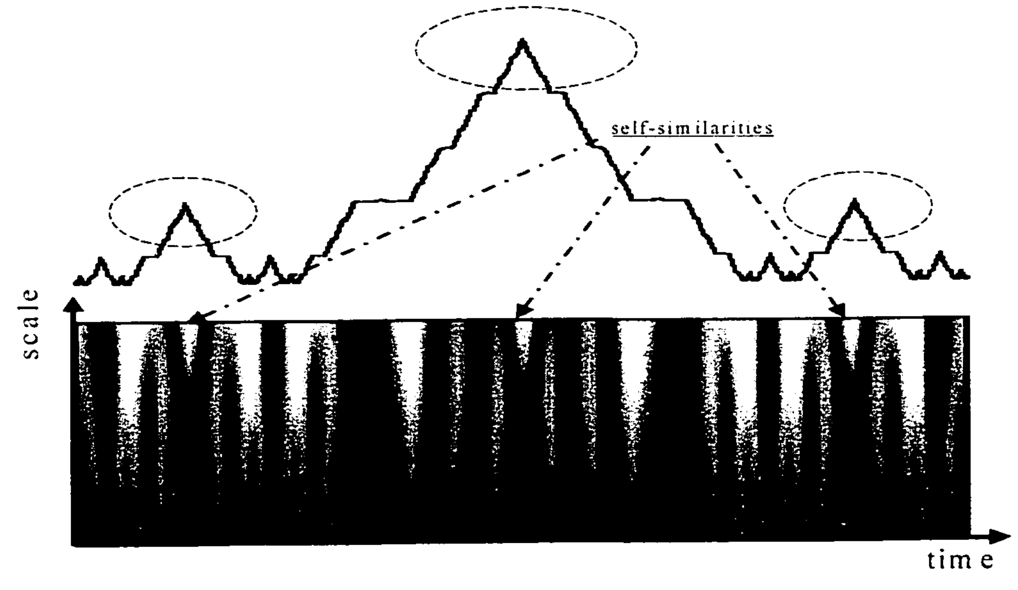


Figure 2-4: Illustration of the self-similarity of the continuous wavelet transform with the VanKoch *fractal signal using the* Ricker *wavelet.*

2.2.2 Inverse transform

For the numerical stability, i.e. the existence of the inverse CWT defined by,

$$f(t) \equiv W^{-1}\{f, \psi\}(a, b) = C_\psi^{-1} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \tilde{f}(t)\psi_{ab}(t)\frac{dbda}{a^2} \tag{2.11}$$

the wavelet $\psi(t)$ must fulfill the (weak) *admissibility condition* given by,

$$0 < C_\psi \equiv \int_{-\infty}^{+\infty} \frac{\left|\hat{\psi}(\omega)\right|^2}{|\omega|} d\omega < \infty \tag{2.12}$$

Elsewhere, if $t\psi(t)$ decays faster than t^{-1}, i.e. $t\psi(t) \in L^2(\mathbb{R})$, then $\hat{\psi}(\omega)$ is continuous and the admissibility condition reduces to the *zero mean condition* [20],

$$\hat{\psi}(0) = 0 \Longleftrightarrow \int_{-\infty}^{+\infty} \psi(t)dt = 0 \tag{2.13}$$

Equation 2.13 says that the wavelet should oscillate (change at least once its sign) and decay at infinity[4].

Remark 2 *In practice, the wavelet is subject to more stringent* decay *and* regularity *conditions. For example, the* Klauder wavelet, *which is generated by the autocorrelation of the* Vibroseis *sweep signal [109], meets the admissibility condition as it is a band-limited signal, with oscillations and decay to zero. However, because of its inherent* ringeness, *it does not decay fast enough. This means a lack of regularity in the frequency domain and thus a bad time-frequency localization [5],[40].*

---

[4] The term "wavelet" (*ondelette* in french) comes precisely form this interpretation.

2.2.3 Properties

The CWT has the following basic properties that it shares with the transform class:

- *Linearity*

$$(\alpha\check{f}_1 + \beta\check{f}_2)(a,b) = \alpha\check{f}_1(a,b) + \beta\check{f}_2(a,b) \tag{2.14}$$

- *Shift-invariance*

$$f_1(t) = f(t-b_1) \Longrightarrow \check{f}_1(a,b) = \check{f}(a,b-b_1) \tag{2.15}$$

- *Scale-invariance*

$$f_1(t) = \frac{1}{\sqrt{\alpha}} f\left(\frac{t}{\alpha}\right) \Longrightarrow \check{f}_1(a,b) = \check{f}\left(\frac{a}{\alpha},\frac{b}{\alpha}\right) \tag{2.16}$$

- *Energy preservation (Plancherel-like formula)*

$$\langle f,f\rangle = \int_{-\infty}^{+\infty} |f(t)|^2\,dt = C_{\psi}^{-1}\int_{0}^{+\infty}\int_{-\infty}^{+\infty} \left|\check{f}(a,b)\right|^2 \frac{dbda}{a^2} \tag{2.17}$$

2.2.4 Examples of wavelets

The two most widely used wavelets with the CWT are the *Ricker* and the *Morlet* wavelets.

- *Ricker wavelet*

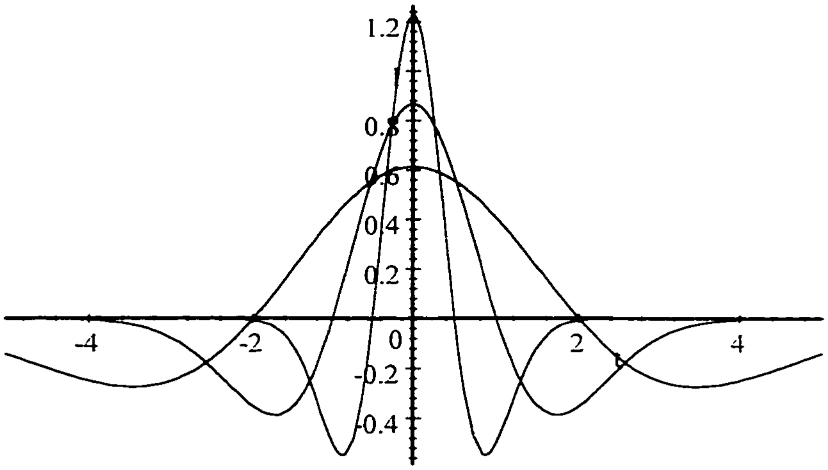


Figure 2-5: Effect of scaling the *Ricker* wavelet $\psi(\frac{t}{a})$ in the time domain.

It is defined as the second derivative of the zero-mean, unit variance *Gaussian* density function[5].

$$\phi(t) = \left(\frac{1}{\pi}\right)^{\frac{1}{4}} e^{-\frac{t^2}{2}} \tag{2.18}$$

that is

$$\psi(t) \equiv \phi''(t) = \frac{2}{\sqrt{3}}\pi^{-\frac{1}{4}}(1-t^2)e^{-\frac{t^2}{2}} \tag{2.19}$$

---

[5] L^2-normalized by the factor $2/\sqrt{3}$.

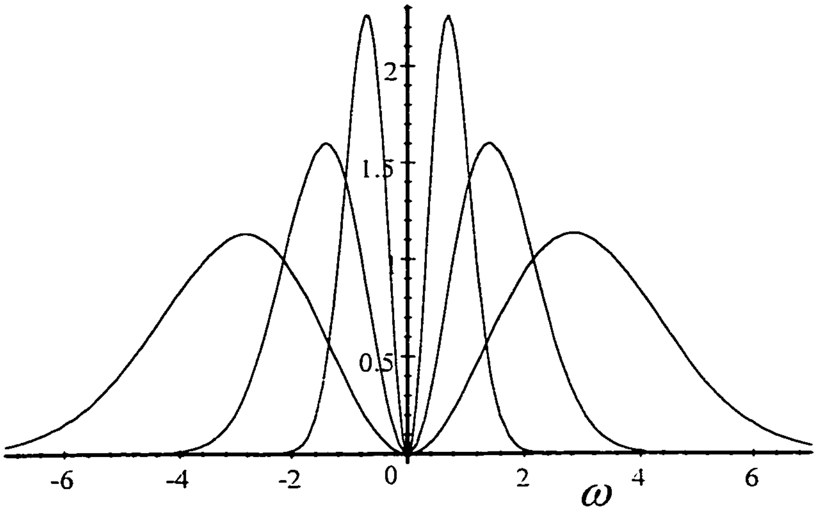


Figure 2-6: Effect of scaling the *Ricker* wavelet $\hat{\psi}(a\omega)$ in the frequency domain.

The effect of scaling the *Ricker* wavelet in both time- and frequency-domains is illustrated in figures 2-5 and 2-6 respectively.

- *Morlet wavelet (Modulated Gaussian)*

It was first used in geophysical exploration by *Grossman* and *Morlet* in 1985 [70] and is at the origin of the development of the wavelet analysis. The wavelet is complex as shown in figure 2-7. It is obtained by shifting a *Gaussian* function in the *Fourier* domain or equivalently multiplying it by an exponential in the time domain. Let ω_0 be the *internal frequency*, then the normalized modulated *Gaussian* density function is defined as

$$\psi(t) \equiv \pi^{-\frac{1}{4}}(e^{-i\omega_0 t} - e^{-\frac{\omega_0^2}{2}})e^{-\frac{t^2}{2}} \tag{2.20}$$

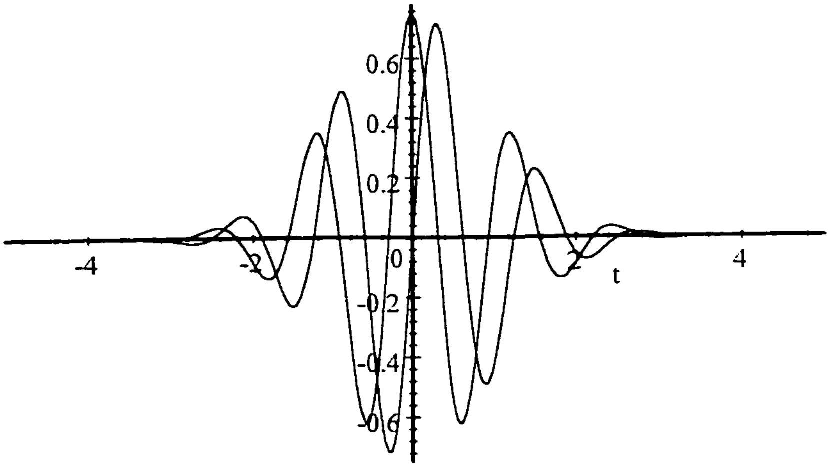


Figure 2-7: *Morlet* wavelet $\psi(t)$ in time domain. The real and imaginary parts have even and odd symmetry respectively.

In practice, it is approximated by[6]

$$\psi(t) \approx \pi^{-\frac{1}{4}} e^{-i\omega_0 t} e^{-\frac{t^2}{2}} \quad \text{for} \quad \omega_0 \approx 5 \tag{2.21}$$

Consequently, the *Fourier* transform of the wavelet that is given by

$$\hat{\psi}(\omega) = \pi^{-\frac{1}{4}} e^{-\frac{(\omega-\omega_0)^2}{2}} \tag{2.22}$$

is slightly adjusted to meet the admissibility condition $\hat{\psi}(0) \approx 0$ as depicted by figure 2-8. Strictly speaking the *Morlet* wavelet is not a wavelet because it is not zero mean. Therefore, the inverse CWT could not be exact [47].

---

[6] In signal processing applications, it has been found that a particularly useful value for ω_0 is the one for which the real part touches the envelope $|\psi|$ at half its height.

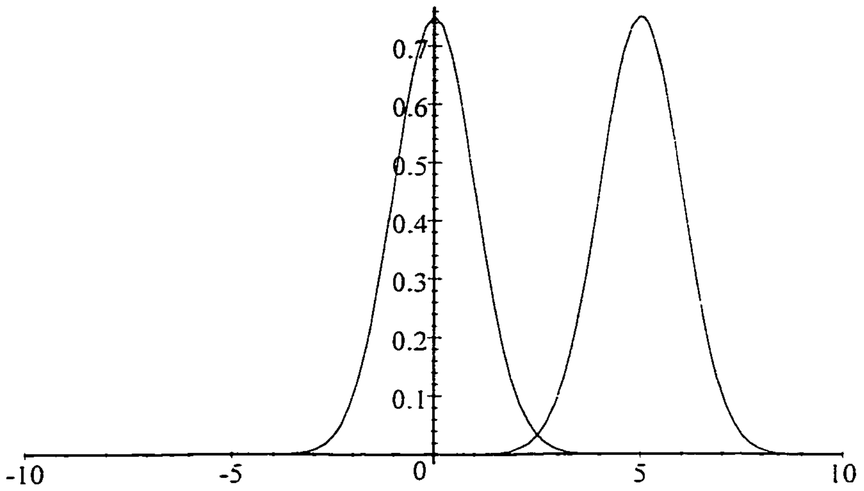


Figure 2-8: Morlet wavelet $\hat{\psi}(\omega)$ in the frequency domain. Note that even with the adjustment $\omega_0 = 5$, the *Fourier* transform does not vanish at the origin but is numerically small.

2.3 Multiresolution Approximation

Multiresolution analysis (or approximation) (MRA) provides a powerful framework to understand, formulate and implement wavelet decompositions using either the *wavelet series transform* (WST) or the *discrete wavelet transform* (DWT). Moreover, the MRA allows for the construction and the design of wavelets associated with *scaling functions.*

2.3.1 Multiresolution theory

The multiresolution theory consists of a nested sequence of closed linear subspaces $\{V_j\}_{j\in\mathbb{Z}}$ in $L^2(\mathbb{R})$ satisfying the following properties:

1. *Monotonicity*[7]

$$\{0\} \subset ... \subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \subset ... \subset L^2(\mathbb{R}) \tag{2.23}$$

2. *Density of the union in* $L^2(R)$

$$\overline{\bigcup_{j \in \mathbb{Z}} V_j} = L^2(\mathbb{R}) \tag{2.24}$$

2. *Trivial intersection*

$$\bigcap_{j \in \mathbb{Z}} V_j = \{0\} \tag{2.25}$$

4. *Scale-invariance*[8]

$$f(t) \in V_j \Longleftrightarrow f(2^j t) \in V_0 \qquad \vee\, j \in \mathbb{Z} \tag{2.26}$$

5. *Shift-invariance*

$$f(t) \in V_0 \Longleftrightarrow f(t-k) \in V_0 \qquad \vee\, j \in \mathbb{Z} \tag{2.27}$$

6. *Existence of a scaling function* $\phi_{0k}(t)$ *such that* $\{\phi(t-k)\}_{k \in \mathbb{Z}}$ *is an orthonormal basis of the central invariant subspace* V_0. *In other terms, the single function* $\phi(t)$

---

[7] We use the convention of decreasing scale indices j , i.e. as j decreases the scale is refined.

[8] Scaling by powers other than two is possible but will not be discussed here.

is the generator of the multiresolution analysis, i.e.

$$V_0 = \overline{span\{\phi(t-n)|n \in \mathbb{Z}} \tag{2.28}$$

2.3.2 Wavelet spaces approximation

As a consequence of the nested scale spaces property (2.23), a signal $f(t)$ can be approximated by successive orthogonal projections $P_{V_j} f$ on the scale spaces V_j as

$$P_{V_j} f = \sum_{k \in \mathbb{Z}} a_{jk} \phi_{jk} = \sum_{k \in \mathbb{Z}} \langle f, \phi_{jk} \rangle \phi_{jk} \tag{2.29}$$

where the *coefficients* a_{jk} are calculated using the inner product as.

$$a_{jk} = \langle f, \phi_{jk} \rangle \equiv \int_{-\infty}^{+\infty} f(t) \phi_{jk}^{*} dt \tag{2.30}$$

Obviously, when going from a scale space V_{j-1} to a coarser space V_j, some precision is lost in the approximation. More precisely, given two consecutive scale spaces $V_j \subset V_{j-1}$, the subspace W_j is defined as the orthogonal complement of V_j in V_{j-1} i.e. $V_j \perp W_j$, such that

$$V_{j-1} = V_j \oplus W_j \tag{2.31}$$

The complementary subspace W_j characterizes those details lost when going from a fine scale space to a coarser one. It is clear from the direct sum (2.31), that the subspaces W_j

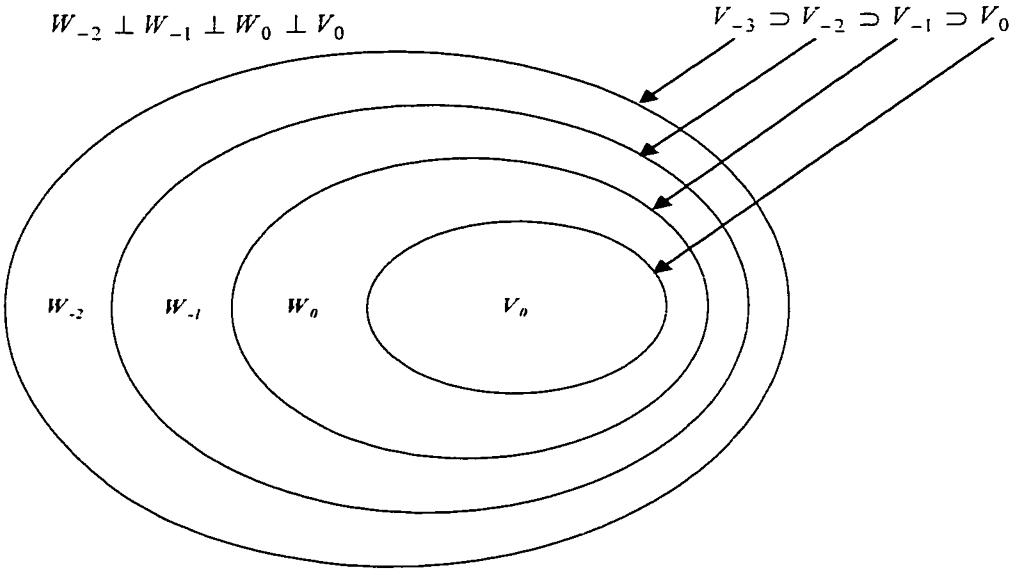


Figure 2-9: Nesting of scaling and wavelet subspaces.

are mutually orthogonal, e.g. $W_j \perp W_{j-1}$. Consequently, the subspaces W_j inherit the scaling property of the scale spaces V_j, i.e.

$$W_j \subset W_{j-1} \tag{2.32}$$

In figure 2-9, the nested scale spaces V_j and their orthogonal complements are depicted for different scales j.

From the nested structure of the scale spaces and the inter-space orthogonality, it follows that if one iterates on the direct sum (2.31) as,

$$\begin{aligned} V_{j-1} &= V_{j+1} \oplus W_{j+1} \oplus W_j \\ &= V_{j+2} \oplus W_{j+2} \oplus W_{j+1} \oplus W_j \end{aligned} \tag{2.33}$$

he gets, for any arbitrary scale index $j > J_0$, where J_0 is the coarsest scale,

$$V_{j-1} = V_{J_0} \oplus \bigoplus_{j=J_0}^{j} W_m \tag{2.34}$$

and for any $j \in \mathbb{Z}$

$$V_{j-1} = \bigoplus_{j=-\infty}^{j} W_m \tag{2.35}$$

and finally, it follows by property (2.24) that

$$L^2(\mathbb{R}) = \overline{\bigoplus_{j \in \mathbb{Z}} W_j} \tag{2.36}$$

Elsewhere, if an orthonormal basis $\{\psi(t-k)\}$ is associated with the central detail subspace W_0, then the family $\psi_{jk}(t) = 2^{-\frac{j}{2}}\psi\left(2^{-j}t - k\right)$ forms an orthonormal basis for the space W_j. The orthonormal basis $\{\psi(t-k)\}$ is referred to as the *wavelet basis*, and the associated detail space the *wavelet space*.

2.3.3 Wavelet series transform

The existence of wavelet bases allows for direct orthogonal projections $P_{W_j}f$ of the signal $f(t)$ on wavelet spaces W_j. This gives a way to directly characterize the details lost when going from a scale space V_{j-1} to a consecutive coarser one V_j. Indeed, using equation

(2.31), one can write

$$P_{W_j} f = P_{V_{j-1}} f - P_{V_j} f \tag{2.37}$$

where the projection on the wavelet spaces is given by

$$P_{W_j} f = \sum_{k \in \mathbb{Z}} d_{jk} \psi_{jk} = \sum_{k \in \mathbb{Z}} \langle f, \psi_{jk} \rangle \, \psi_{jk} \tag{2.38}$$

Again, the coefficients d_{jk} are calculated using the inner product as

$$d_{jk} = \langle f, \psi_{jk} \rangle \equiv \int_{-\infty}^{+\infty} f(t) \psi_{jk}^{*} dt \tag{2.39}$$

In the limit case given by equation (2.36), projection (2.38) yields a discrete reconstruction formula of the signal from its projections on the wavelet spaces W_j, i.e.

$$f(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} d_{jk} \psi_{jk} \tag{2.40}$$

The double indexed sum (2.40) is referred to as the *wavelet series transform* (WST)[9].

2.3.4 Finite wavelet series transform

At the first glance, the wavelet expansion (2.40) seems to have the same problems in calculation that an infinite *Fourier* series has. However, in the practical situation, the summation over the scale index j and the summation over the translation index k can

---

[9] or *wavelet expansion.*

be made finite with little errors for some norm. Firstly, recall that in practice we do not process continuous-time signals but rather discrete-time signals. Thus the finest scale is not $L^2(\mathbb{R})$, but the sample space $V_{J_1}, 0 >> J_1 > -\infty$, which consists of the orthogonal projection[10] of $f(t)$ onto V_{J_1} such that $\left\| f - P_{V_{J_1}} f \right\| \simeq 0$. Note that this is not a restriction because most of the real life signals have some natural scale associated with them due to the non-vanishing support of the sensors. The WST (2.40) becomes then[11]

$$f(t) \approx P_{V_{J_1}} f = \sum_{j=-\infty}^{J_1+1} \sum_{k=-\infty}^{\infty} d_{jk} \psi_{jk} \tag{2.41}$$

Secondly, the lower limit on the scale index j is made finite by introducing the scaling function $\phi_{J_0 k}$ at the coarsest scale space V_{J_0}, at which the signal features are slowly changing or are of longer duration, i.e. contain low frequencies. This yields,

$$f(t) \approx P_{V_{J_1}} f = \sum_{k=-\infty}^{\infty} a_{J_0 k} \phi_{J_0 k} + \sum_{j=J_0}^{J_1+1} \sum_{k=-\infty}^{\infty} d_{jk} \psi_{jk} \tag{2.42}$$

where $a_{J_0 k}$ are the residual coefficients. Notice that the negative index j is absorbed by the scaling function $\phi_{J_0 k}$ at the coarsest scale J_0 according to

$$V_{J_0} = W_{J_0+1} \oplus W_{J_0+2} \oplus ... \oplus W_0 \oplus ... \oplus W_{-\infty} \tag{2.43}$$

---

[10] The Shannon sampling is one instance of such a projection.

[11] Recall that we use the decreasing index convention, i.e. $V_{j+1} \subset V_j$ or V_j is finer than V_{j+1}.

leading to the following underlying subspaces decomposition of the approximation (2.42)

$$V_{J_1} = V_{J_0} \oplus W_{J_0} \oplus W_{J_0-1} \oplus ... \oplus W_{J_1+2} \oplus W_{J_1+1} \tag{2.44}$$

Finally if the signal $f(t)$ is of finite duration and if both the scaling and the wavelet functions have a finite support, then the shift index k is also made finite. We get then the following expression for the *finite* WST

$$f(t) \approx P_{V_{J_1}} f = \sum_{k \in \mathbb{Z}} a_{J_0 k} \phi_{J_0 k} + \sum_{j=J_0}^{J_1+1} \sum_{k \in \mathbb{Z}} d_{jk} \psi_{jk} \tag{2.45}$$

2.4 Discrete Wavelet Transform

2.4.1 Wavelet decomposition and reconstruction

The finite WST (2.45) can be interpreted as a sequence of *recursive* orthogonal projections of the signal $f(t)$ onto wavelet spaces W_j up to the coarsest scale space V_{J_0}. According to the space decomposition (2.44), one can write

$$P_{V_{J1}} f = P_{V_{J_0}} f + P_{W_{J_0}} f + P_{W_{J_0-1}} f + ... + P_{W_{J_1+2}} f + P_{W_{J_1+1}} f \tag{2.46}$$

or symbolically

$$f^{J_1} \equiv f^{J_0} + g^{J_0} + g^{J_0-1} + ... + g^{J_1+2} + g^{J_1+1} \tag{2.47}$$

This recursive process is illustrated schematically by the diagram of figure 2-10. The

$$\begin{tikzcd}
f^{J_1} \arrow[r] \arrow[dr] & f^{J_1+1} \arrow[r] \arrow[dr] & \cdots \arrow[r] \arrow[dr] & f^{J_0-1} \arrow[r] \arrow[dr] & f^{J_0} \\
 & g^{J_1+1} & \cdots & g^{J_0-1} & g^{J_0}
\end{tikzcd}$$

Figure 2-10: Wavelet multiresolution decomposition of a signal f.

$$\begin{tikzcd}
f^{J_0} \arrow[r] & f^{J_0-1} \arrow[r] & \cdots \arrow[r] & f^{J_1+1} \arrow[r] & f^{J_1} \\
g^{J_0} \arrow[ur] & g^{J_0-1} \arrow[ur] & \cdots \arrow[ur] & g^{J_1+1} \arrow[ur] &
\end{tikzcd}$$

Figure 2-11: Wavelet multiresolution reconstruction of a signal f.

wavelet decomposition gives a *multiresolution (or multiscale) analysis* of a signal $f(t)$ in terms of its orthogonal projections g_j onto the wavelet spaces W_j and the residual f_{J_0}, which consists of the projection of the signal $f(t)$ on the coarsest scale space V_{J_0}.

Remark 3 *Since $W_l \perp W_m$ if $l \neq m$ and $V_j \perp W_j$, the wavelet multiresolution decomposition is unique once the spaces V_j and W_j are selected.*

It is also desirable to reconstruct the signal $f(t)$ from its wavelet representation using a recursive process similar to the decomposition of figure 2-10. Since $W_j \subset V_{j-1}$ and $V_j \subset V_{j-1}$, the original signal can be obtained from its projections. The wavelet reconstruction is thus essentially the reverse of the decomposition, as illustrated by figure 2-11.

2.4.2 Approximation and details wavelet coefficients

To implement the wavelet decomposition, the projections onto the different spaces V_j and W_j need to be computed. Recall that the two sets of functions $\{\phi_{jk}\}$ and $\{\psi_{jk}\}$ defined by

$$\phi_{jk} = 2^{-\frac{j}{2}}\phi(2^{-j} - k) \tag{2.48}$$

and

$$\psi_{jk} = 2^{-\frac{j}{2}}\psi(2^{-j} - k) \tag{2.49}$$

are orthogonal bases for $\{V_j, W_j\} \subset L^2(\mathbb{R})$. Consequently, the coefficients of the orthogonal projections, given by equations (2.30) and (2.39), are respectively expressed as

$$a_{jk} = \langle f, \phi_{jk} \rangle = 2^{-\frac{j}{2}} \int f(t)\phi^*(2^{-j}t - k)dt \tag{2.50}$$

$$d_{jk} = \langle f, \psi_{jk} \rangle = 2^{-\frac{j}{2}} \int f(t)\psi^*(2^{-j}t - k)dt \tag{2.51}$$

The set of coefficients $\{a_{jk}, d_{jk}\}_{jk\in\mathbb{Z}}$ given by equations (2.50) and (2.51) is called the *discrete wavelet transform* (DWT) of the signal $f(t)$. The coefficients a_{jk}, which convey the low frequency trends of the signal $f(t)$ are called the approximation coefficients. The detail coefficients d_{jk}, as their part, bear the high-frequency trends.

In order to compute the wavelet coefficients without computing the integrals, it is judicious to take advantage of the fact that both the multiresolution decomposition and reconstruction require only projections between consecutive subspaces of the multiresolution ladder (2.23). This leads to the *two-scale difference equation.*

2.4.3 Two-scale difference equation

The scaling property (2.26) expressed using two consecutive subspaces leads to the so-called two-scale difference equation. The latter characterizes the scaling and wavelet functions through a sequence of *filter coefficients*. Next, the derivation of the two-scale difference equation is performed in both time- and frequency-domains.

Time-domain characterization

Consider the scaling function $\phi \equiv \phi_{0k}(t) = \phi(t-k) \in V_0 \subset V_{-1}$, where the finer subspace is endowed with the scaling orthogonal basis $\phi_{-1} \equiv \phi_{-1k}(t) = 2^{\frac{1}{2}}\phi(2t-k)$. The scaling function ϕ can be expressed as a linear combination

$$\phi = \sum_{k\in\mathbb{Z}} c_k \phi_{-1k} = \sqrt{2}\sum_{k\in\mathbb{Z}} c_k \phi(2t-k) \tag{2.52}$$

where the *scaling filter coefficients* c_k, that unambiguously characterize the scaling function $\phi(t)$, are given by

$$c_k = \langle \phi_{0k}, \phi_{-1k} \rangle \tag{2.53}$$

Similarly, given a wavelet $\psi \equiv \psi_{0k}(t) = \psi(t-k) \in W_0 \subset V_{-1}$ and some real coefficients b_k one can write

$$\psi = \sum_{k \in \mathbb{Z}} b_k \phi_{-1k} = \sqrt{2} \sum_{k \in \mathbb{Z}} b_k \phi(2t-k) \tag{2.54}$$

where the *wavelet filter coefficients* are given by

$$b_k = \langle \psi_{0k}, \phi_{-1k} \rangle \tag{2.55}$$

Elsewhere, using the fact that $\psi(t)$ and $\phi(t)$ are mutually orthogonal, i.e.

$$\int_{-\infty}^{+\infty} \phi(t-n)\phi(t-m)dt = 0 \tag{2.56}$$

the wavelet function can be expressed using the scaling filter coefficients as [20],

$$\psi(t) = \sum_{k \in \mathbb{Z}} (-1)^k c_{N-1-k} \phi(2t-k) \tag{2.57}$$

or equivalently,

$$b_k \equiv (-1)^{k+1} c_{N-1-k} \tag{2.58}$$

where N is an arbitrary integer[12] chosen to conveniently position the wavelet filter coefficients.

---

[12] In fact, N is the number of nonzero coefficients c_k.

Remark 4 *Sometimes, the set of scaling filter coefficients* $h_k = \sqrt{2}c_k$ *and wavelet filter coefficients* $g_k = \sqrt{2}c_k$ *is used.*

Frequency-domain characterization

The wavelet transform and MRA are much easier analyzed and studied in the *Fourier* domain because most of the encountered identities, which have a convolutional structure, become multiplicative identities. In addition, the design of wavelet filters is in general performed in the frequency domain.

Applying the *Fourier* transform to both sides of the equality (2.52) yields

$$\begin{aligned} \hat{\phi}(\omega) &= \sqrt{2}\sum_{k\in\mathbb{Z}} c_k \int_{-\infty}^{+\infty} \phi(2t-k)e^{i\omega t}dt \\ &= m_0\left(\frac{\omega}{2}\right)\hat{\phi}\left(\frac{\omega}{2}\right) \end{aligned} \tag{2.59}$$

The trigonometric quantity

$$m_0(\omega) \equiv \frac{1}{\sqrt{2}}\sum_{k\in\mathbb{Z}} c_k e^{i\omega k} \tag{2.60}$$

is the crucial function in the theory of multiresolution analysis. It is 2π-periodic and $m_0 \in L^2([0,2\pi])$ because $\sum_{k\in\mathbb{Z}} |c_k|^2 < \infty$. Note also that as $\hat{\phi}(0) = 1$, then $m_0(0) = 1$ [20].

Applying equation (2.59) recursively for $\omega/2, \omega/4, \ldots$ and recalling that[13] $\hat{\phi}(0) = \int_{-\infty}^{+\infty} \phi(t)dt = 1$, one gets in the limit case, the following expression of the two-scale difference equation

---

[13] If the Fourier integral is defined as $\phi(\omega) = \frac{1}{\sqrt{2\pi}}\int \phi(t)e^{-i\omega t}dt$ then $\hat{\phi}(0) = \frac{1}{\sqrt{2\pi}}$.

in the *Fourier* domain

$$\hat{\phi}(\omega) = \prod_{j=1}^{\infty} m_0\left(\frac{\omega}{2^j}\right) \tag{2.61}$$

Finally, note that by applying a similar procedure, equation (2.54) can be expressed in the frequency domain as

$$\hat{\psi}(\omega) = m_1\left(\frac{\omega}{2}\right)\hat{\phi}\left(\frac{\omega}{2}\right) \tag{2.62}$$

where

$$m_1(\omega) \equiv \frac{1}{\sqrt{2}}\sum_{k\in\mathbb{Z}} b_k e^{i\omega k} \tag{2.63}$$

Moreover, the *Fourier* transform of the wavelet (2.62) can be expressed in terms of the polynomial $m_0(\omega)$ up to a periodic function (i.e. not uniquely) as [20]

$$\hat{\psi}(\omega) = e^{-\frac{i\omega}{2}} m_0^*\left(\frac{\omega}{2}+\pi\right)\hat{\phi}\left(\frac{\omega}{2}\right) \tag{2.64}$$

2.5 *Haar* Multiresolution Analysis

For the sake of illustration, the *Haar* wavelet system is derived. For, consider the box function

$$\phi(t) = \chi_{[0,1]}$$

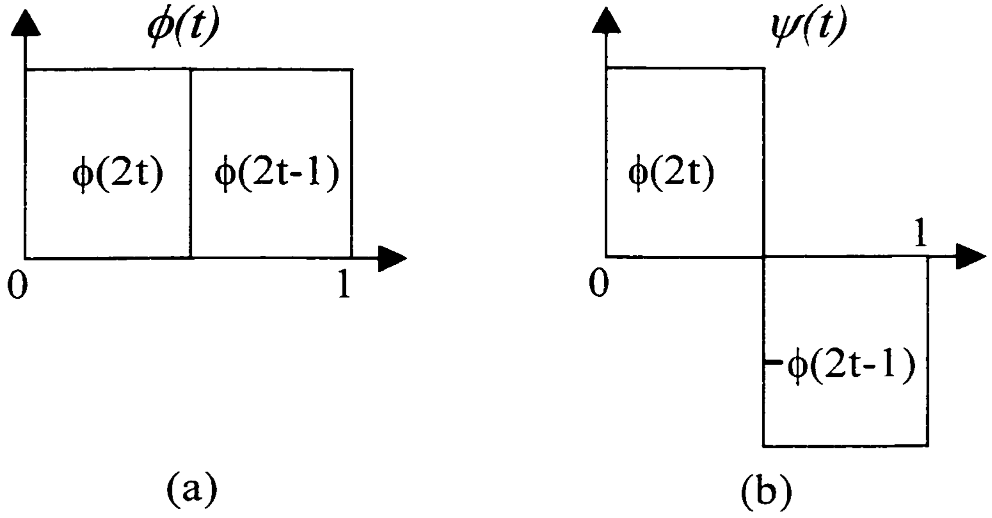


Figure 2-12: *Haar* scaling function and wavelet in the domain.

which has compact support [0, 1] and whose *Fourier* transform $\hat{\phi}(\omega) = \frac{\sin \pi\omega}{\pi\omega}$ (which has an infinite support!)[14] defines a low-pass filter. Then from the two-scaling difference equation, one gets

$$h_k = \int_0^1 \phi(t)\phi(2t-k)dt$$

It is easy to verify that only $\phi(t)$ and $\phi(2t-1)$ are not disjoint from $\phi(t)$, and thus the non-null coefficients are only $h_0 = h_1 = \frac{1}{2}$. Consequently, the two-scale equation illustrated by figure 2-12-(a), can be written as

$$\phi(t) = \phi(2t) + \phi(2t-1)$$

---

[14] Note that the *Heisenberg* inequality is violated.

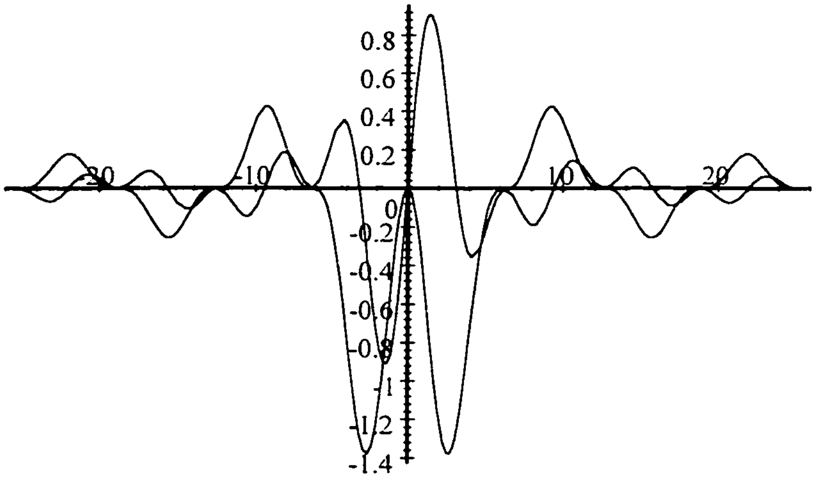


Figure 2-13: *Haar* wavelet in frequency domain. The real and imaginary parts have even and odd symmetry respectively.

Using $g_k \equiv (-1)^{k+1} h_{N-1-k}$, with $N = 2$ and $k = 0, 1$, we get $g_0 = -h_1 = -\frac{1}{2}$ and $g_1 = h_0 = \frac{1}{2}$. Thus the wavelet is given according to equation (2.57) as,

$$\psi(t) = \phi(2t) - \phi(2t - 1)$$

which gives the *Haar* wavelet displayed in figure 2-12-(b). In the frequency domain, the *Haar* wavelet is given by [47]

$$\hat{\psi}(\omega) = 2\frac{1 - \cos\omega}{\omega} e^{\frac{1}{2}i(\omega+\pi)}$$

and is plotted in figure 2-13.

Notice that due to the poor time-domain regularity of the *Haar* wavelet, the oscillations decrease slowly in the frequency domain.

The *Haar* wavelet was actually the first wavelet to be used in mathematics. It was introduced by *A. Haar* in 1906 [70], who also proved that the set $\{2^{-\frac{j}{2}}\psi(2^{-j}t-k)\}_{j,k\in\mathbb{Z}}$, is an orthonormal basis of $L^2(\mathbb{R})$. Furthermore, the *Haar* wavelet is the only known wavelet that is orthogonal, symmetric and compactly supported [20]. However, its discontinuity prevents from using it in practice. In the next chapter, it will be shown how various families of wavelets with many interesting properties can be designed using the MRA framework.

Chapter 3

WAVELETS DESIGN

The performance of the DWT relies on both the choice of the wavelet bases and the algorithms to implement it. This chapter is concerned with the former aspect, whereas the latter aspect will be covered in details in the next two chapters. Wavelets systems can be either *orthogonal* or *non-orthogonal*. An orthogonal MRA system is one in which the scaling functions are orthogonal to one another, the wavelets are orthogonal to one another, and each wavelet is orthogonal to every coarser scaling function. The most famous orthogonal wavelets are the *compactly supported Daubechies* wavelets [19]. Orthogonality leads to a robust and symmetric implementation using filter banks. It places however, strong limitations on the possibilities of the system. Indeed, there exist no wavelets (besides the *Haar* wavelets) that are at once orthogonal, symmetric and compactly supported [92]. To improve an orthogonal system, orthogonality should be sacrificed either partially or completely. This yields the *semi-orthogonal* and the *biorthogonal* wavelet systems respectively. In the semiorthogonal MRA, the wavelets[1] at a given resolution are at least orthogonal to all coarser scaling functions, though not necessarily orthogonal to each other. Unfortunately, the semiorthogonal wavelet systems suffer from some limitations when extended to higher dimensions, which is not the case for the biorthogonal wavelet systems [68]. The latter use two dual MRA such that the primal scaling functions are orthogonal to the dual wavelets and the primal wavelets are orthogonal to the dual scaling functions. Moreover, from the implementation point of view, any biorthogonal system can be factored using the *lifting scheme* into elementary cells leading to the most efficient computation of the DWT. The *Cohen-Daubechies-Feauveau* (CDF) class [17] forms the most famous biorthogonal wavelets. In principle, wavelet systems can be constructed by seeking a smooth time function which fulfills the MRA requirements. This approach does not lend itself to a systematic design of large wavelet families. Elsewhere, it is only restricted to the semiorthogonal and orthogonal instances. Moreover the constructed orthogonal wavelet systems do not have a compact support. Indeed, the wavelet support is controlled by the decay of the wavelet in either the frequency domain, as it is the case for the *Meyer* wavelets [70] which decay as ω^{-1} or in the time domain, as it is the case for the *Battle-Lemarié* spline wavelets [53] that have an exponential decay. Finally, the regularity and the symmetry of the wavelets cannot be explicitly specified in the construction stage.

On the contrary, the design of wavelets families using the framework of MRA is very efficient as it uses the connection of wavelets with digital filters. This method has been recently initiated by *Daubechies* [17],[19], and leads to the design of complete families of

---

[1] Also called pre-wavelets.

orthogonal and biorthogonal wavelets systems endowed with many appealing properties such that regularity, orthogonality, symmetry, and compactness of support. Firstly, the chief properties of wavelets will be presented and their role in data compression will be highlighted. Secondly, these rather theoretical properties will be translated into more practical design constraints in terms of the scaling filter coefficients. Indeed, these constitute the hub for the design of wavelet families. Thirdly, the design of orthogonal and biorthogonal wavelets will be discussed. Lastly, the extension of MRA to the two-dimensional case will be discussed.

3.1 Wavelet Properties

The *Haar* wavelet presented in section 2.5 is the first and sole wavelet that is orthogonal, symmetric and compactly support on the interval $[0, 1]$. However, its discontinuity prevents it from being used in most of the applications. Since then, a variety of new wavelets with excellent properties, have been designed. In the sequel, the wavelet properties[2] that are of interest for data compression will be reviewed.

3.1.1 Vanishing moments versus size of support

Vanishing moments - The magic of wavelets rely on their ability to efficiently approximate particular classes of signals $f(t)$ with few non-zero wavelet coefficients $\langle f, \psi_{jk} \rangle$.

---

[2] These wavelets inherits also some of the properties characterizing wavelets developed for the CWT, e.g. time-frequency localization.

The degree of approximation is controlled by the number of *vanishing moments*, i.e.

$$\int_{-\infty}^{\infty} t^m \psi(t) dt = 0, \quad m = 0, 1, ..., p - 1 \tag{3.1}$$

Stated otherwise, the *decorrelation power* and *energy compaction* properties allow the concentration of most of the signal energy in few wavelet coefficients. A further compression[3] occurs when only the most significant coefficients are kept and all the rest killed. For example, if the signal is modeled by a constant polynomial then the *Haar* wavelet, that has one vanishing moment[4], will lead to an optimal compact decomposition of the signal with only the approximation coefficients a_k, as all the wavelet coefficients d_k are nil. In sum, the higher the number of vanishing moments is, the better smooth signals can be approximated with few wavelet coefficients.

Size of support - If the scaling filter coefficients c_k have a finite support then the scaling function $\phi(t)$ has the same finite support. Moreover, if the support of $\phi(t)$ is $[N_1, N_2]$ (or $[0, 2p - 1]$) then the support of $\psi(t)$ is $[\frac{N_1 - N_2 + 1}{2}, \frac{N_2 - N_1 + 1}{2}]$ (or $[-p + 1, p]$).

A priori, the support size and the number of vanishing moments are independent. However the orthogonality constraint on the wavelets, i.e. $\langle \psi_{jk}, \psi_{jl} \rangle = 0$, implies that if the latter have p vanishing moments then their support is at least $2p - 1$. Thus, the adequate selection of a particular wavelet is subject to a trade-off between the size of the support and the number of vanishing moments. As a guide line, if $f(t)$ has few

---

[3] To be distinguished from the equally challenging problem of quantization and encoding dealt with in chapter 6.

[4] which is the minimum requirement and is nothing but the admissiblity condition (2.13).

isolated singularities and is very regular elsewhere, a wavelet with many vanishing moments (and thus of large support size) is desirable to produce a large number of small wavelet coefficients d_k because the smooth portion of the signal is large. However, if the density of the singularities increases, it might be better to decrease the size of the support of ψ at the cost of reducing the number of vanishing moments, because wavelets that overlap the singularities create high amplitude coefficients, which may increase the dynamic range of the wavelet coefficients [68].

3.1.2 Regularity

The importance of the regularity or smoothness[5] of wavelets appears in many respects. Firstly, the regularity has mostly a cosmetic effect on the error introduced by wavelet coefficients processing, e.g. quantization. When reconstructing a signal from its wavelet coefficients, an error ϵ added to a coefficient $d_k = \langle f, \psi_{jk} \rangle$ will add the wavelet component $\epsilon\psi_{jk}$ to the reconstructed signal $\check{f}(t)$. If $\psi(t)$ is smooth, then the error will also be smooth and leads to less visible artifacts than in the case of irregular errors[6]. Secondly, wavelets are good tools to study either the local or global regularity of signals. Lastly, the regularity of discrete wavelets is important to assess the convergence of iterated filter bank that is used to implement the DWT.

---

[5] Stricly speaking, the regularity concerns the wavelet filter and the smoothness concerns the wavelet function.

[6] i.e. with the same L^2 but different L^1 norms.

3.1.3 Symmetry

Usually, with symmetric wavelet and scaling functions, fewer artifacts appear in the reconstructed data. This is very important as far as the visual aspect of the data is of concern. Symmetric functions also simplify the handling of boundaries as it will explained in chapter 4.

3.1.4 Rational coefficients

For computer implementation it is preferable that the scaling and wavelet filters coefficients are rationals or even better, dyadic rationals. Indeed, a binary multiplication by a power of two translates to bits shifts.

3.2 Orthogonal Wavelet System

3.2.1 Scaling filter constraints [12],[20]

Normalization condition

If $\phi(t)$ is a solution to the two-scale difference equation and if $\int \phi(t)dt = 1$ then

$$\sum_{k\in\mathbb{Z}} c_k = \sqrt{2} \tag{3.2}$$

Furthermore, this normalization condition implies that the frequency response of the *finite impulse response* (FIR) scaling filter at DC ($\omega = 0$) is $\sqrt{2}$.

Orthogonality condition

The orthogonality of the scaling functions given by

$$\int_{-\infty}^{+\infty} \phi(t)\phi^*(t-n)dt = \delta(n) \tag{3.3}$$

can be expressed in terms of the scaling filter coefficients as

$$\sum_{k\in\mathbb{Z}} c_k c_{k-2n} = \delta(n) \tag{3.4}$$

Using the *Parseval* identity, an equivalent condition can be developed in the frequency domain. Indeed, equation (3.3) can be written as,

$$\delta(n) = \int_{-\infty}^{+\infty} \left|\hat{\phi}(\omega)\right|^2 e^{in\omega} d\omega \tag{3.5}$$

which becomes, by virtue of the *Poisson* formula,

$$\delta(n) = \int_0^{2\pi} \sum_{q\in\mathbb{Z}} \left|\hat{\phi}(\omega+2q\pi)\right|^2 d\omega \tag{3.6}$$

The above equation implies that

$$\sum_{q\in\mathbb{Z}} \left|\hat{\phi}(\omega+2q\pi)\right|^2 = \frac{1}{2\pi} \tag{3.7}$$

which by substituting for $\hat{\phi}$ using equation (2.59), yields

$$\sum_{q\in\mathbb{Z}} |m_0(\omega+q\pi)|^2 \left|\hat{\phi}(\omega+q\pi)\right|^2 = \frac{1}{2\pi} \tag{3.8}$$

Furthermore, since m_0 is 2-π periodic, one can further split the above sum into even and odd components as,

$$|m_0(\omega)|^2 \sum_{q\in\mathbb{Z}} \left|\hat{\phi}(\omega+2q\pi)\right|^2 + |m_0(\omega+\pi)|^2 \sum_{q\in\mathbb{Z}} \left|\hat{\phi}(\omega+(2q+1)\,\pi)\right|^2 = \frac{1}{2\pi} \tag{3.9}$$

which finally gives the expression of the orthogonality condition in the frequency domain

$$|m_0(\omega)|^2 + |m_0(\omega+\pi)|^2 = 1 \tag{3.10}$$

Vanishing moments conditions

In practice, condition (3.1) is also replaced by a set of constraints on the scaling filter coefficients known as *sum rules* [21], i.e.

$$\sqrt{2}\sum_{k\in\mathbb{Z}}(-1)^k k^m c_k = 0, \qquad m = 0, 1, ..., p-1 \tag{3.11}$$

The sum rules condition is equivalent to having a zero of order p at $\omega = \pi$ for the fundamental trigonometric polynomial[7] $m_0\left(e^{i\omega}\right) \equiv \frac{1}{\sqrt{2}} \sum_{k\in\mathbb{Z}} c_k e^{i\omega k}$, i.e.

$$m_0\left(e^{i\omega}\right) = \left(\frac{1+e^{i\omega}}{2}\right)^p R_p\left(e^{i\omega}\right) \tag{3.12}$$

where $p \succeq 1$. Factorization (3.12) gives a sufficient condition on the regularity of iterated scaling filters, i.e. the convergence of the scaling function $\phi(t)$ associated with the scaling filter to a smooth time function. Note by passing that $m_0^{(m)}(\pi)$ also yields condition (3.11).

Example 1 ***(Daubechies 4-coefficient wavelet)***

At this stage, we have all the ingredients to design the famous Daubechies' *wavelet with two vanishing moments and support size* $[0,3]$. *We need four equations, but we effectively have five equations. For calculation simplicity, we only select linear equations. From the normalization equation (3.2), we get*

$$c_0 + c_1 + c_2 + c_3 = \sqrt{2} \tag{i}$$

The orthogonality condition (3.4) generates two additional equations, i.e.

$$c_0^2 + c_1^2 + c_2^2 + c_3^2 = 1 \tag{ii}$$

$$c_0 c_2 + c_1 c_3 = 0 \quad \mathit{for} \quad n = \pm 1 \tag{iii}$$

---

[7] Either convenient, the frequency argument will be either ω or $e^{i\omega}$.

and finally the sum rules condition (3.11) yields the last two equations. i.e.

$$c_0 - c_1 + c_2 - c_3 = 0 \quad \mathit{for} \quad m = 0 \tag{iv}$$

$$-c_1 + 2c_2 - 3c_3 = 0 \quad \mathit{for} \quad m = 1 \tag{v}$$

From the equations $(i), (ii), (iv)$ *and* (v), *one gets*

$$\begin{aligned} c_0 &= \tfrac{1\pm\sqrt{3}}{4\sqrt{2}}; \quad c_1 = \tfrac{3\pm\sqrt{3}}{4\sqrt{2}} \\ c_2 &= \tfrac{3\mp\sqrt{3}}{4\sqrt{2}}; \quad c_3 = \tfrac{1\mp\sqrt{3}}{4\sqrt{2}} \end{aligned} \tag{vi}$$

Clearly, one set of solutions is the antithesis of the other, i.e. if one leads to $\phi(t)$ *then the other will lead to* $\phi(-t)$. *Both satisfy the conditions of normality, orthogonality and exact representation of a linear signal because of the two vanishing moments. The wavelet filter follows directly from the relationship (2.58).*□

In principle, by assigning different numbers of vanishing moments, a similar procedure can be used to design a whole family of compactly supported orthogonal wavelets. However, the design is better performed in the frequency domain to take advantage of the mature subject of digital filtering for deriving some interesting variants as it will be discussed in the sequel.

3.2.2 Compactly supported orthogonal wavelets

The construction of wavelets as initiated by *Daubechies* [19], [20] is performed in the frequency domain because translation and dilation become algebraic equations. It consists essentially of three stages, namely the *algebraic*, the *analytic*, and the *geometric* stages. Throughout these stages, the four following properties should be assured [95]:

1. Wavelets form a *Riesz* basis for $L^2(\mathbb{R})$.
2. The wavelets are either orthogonal or the dual (biorthogonal) wavelets are known.
3. The wavelets and their duals are local in space and frequency, and even compactly supported. The frequency localization follows from the smoothness of the wavelets (*decay towards high frequencies*) and the fact that they have vanishing (polynomial) moments (*decay towards low frequencies*).
4. Wavelets fit into the framework of MRA to lead to the FWT. The later allows to toggle between the function and it wavelet coefficients in linear time $\mathcal{O}(N)$.

The *algebraic stage* concerns the design of the scaling filter and assures that *property 4* is satisfied. In the *analytic stage*, it is shown that (time) functions[8] associated with these filters exist (*property 2*), that they are localized (*property 3*), and that they form a basis for the proper function space (*property 1*). Finally, in the *geometric stage*, the smoothness of the basis functions is checked[9] (*property 3*).

---

[8] i.e. scaling functions and wavelets.

[9] Note that the functions $\psi(t)$ and $\phi(t)$ does intervene explicitly in the implementation of the DWT!

Daubechies [20] designed three families of compactly supported orthogonal wavelets, namely the *most-asymmetric* wavelets or *daublets*, the *least-asymmetric* wavelets or *symmlets*, and the *coiflets*[10].

Most-asymmetric wavelets

The daublets are designed to have the minimum compact support $[2p-1]$ for an arbitrary regularity, expressed in terms of the number of vanishing moments p. In the discrete filter design parlance, the scaling filter $m_0(e^{i\omega})$ should have p zeros at the *Nyquist* frequency π, i.e.

$$m_0(e^{i\omega}) = \left(\frac{1+e^{i\omega}}{2}\right)^p \mathcal{R}_p(e^{i\omega}) \tag{3.13}$$

Such a discrete filter is called *p-regular* or *maxflat* because the frequency response is maximally flat at $\omega = 0$ and $\omega = \pi$ [92]. The associated wavelet function defined by equation (2.64) will automatically have p vanishing moments and will be able to approximate polynomials up to order $p-1$. The remaining p degrees of freedom are exploited to ensure the orthogonality condition (3.10) through the design of the residual 2-π periodic trigonometric polynomial $\mathcal{R}_p(e^{i\omega})$.

---

[10] named in honour of *R. Coifman*.

Let the *reciprocal product polynomial* $P(e^{i\omega})$ be

$$\begin{aligned} P(e^{i\omega}) &\equiv \left|m_0(e^{i\omega})\right|^2 \\ &= m_0(e^{i\omega})m_0^*(e^{i\omega}) \\ &= \left(\frac{1+e^{i\omega}}{2}\right)^p \left(\frac{1+e^{-i\omega}}{2}\right)^p \left|\mathcal{R}_p(e^{i\omega})\right|^2 \\ &= \left(\cos^2\frac{\omega}{2}\right)^p R_p\left(\sin^2\frac{\omega}{2}\right) \end{aligned} \tag{3.14}$$

where the trigonometric polynomial $R_p(\omega) = |\mathcal{R}_p(\omega)|^2$ is written as a polynomial in $\sin^2\frac{\omega}{2}$. The orthogonality constraint (3.10) becomes then

$$(1-y)^p R_p(y) + y^p R_p(1-y) = 1 \tag{3.15}$$

where $y = \sin^2\frac{\omega}{2}$ for all $y \in [0,1]$. The solution of equation (3.15) is obtained by applying the following theorem to the case where $p_1(y) = (1-y)^p$ and $p_2(y) = y^p$.

Theorem 2 ***(Bezout)*** *If p_1 and p_2 are two polynomials with no common zeros, then there exist unique polynomials q_1 and q_2 such that*

$$p_1(y)q_1(y) + p_2(y)q_2(y) = 1$$

where p_1, p_2, q_1, q_2, are of degree $n_1, n_2, n_2 - 1, n_1 - 1$, respectively. □

Therefore, one gets

$$(1-y)^p q_1(y) + y^p q_2(y) = 1 \tag{3.16}$$

and by substituting $1-y$ for y

$$(1-y)^p\, q_2(1-y) + y^p q_1(1-y) = 1 \tag{3.17}$$

As q_1 and q_2 are unique, then $q_1(y) = q_2(1-y)$ and $q_2(y) = q_1(1-y)$. This yields the following explicit expression for $R_p(y) = q_1(y)$ [20],

$$\begin{aligned} R_p(y) &= (1-y)^{-p}[1 - y^p q_1(1-y)] \\ &= \sum_{k=0}^{p-1} \begin{pmatrix} p+k-1 \\ k \end{pmatrix} y^k + \mathcal{O}(y) \end{aligned} \tag{3.18}$$

where we expanded the first p terms of the *Taylor* series for $(1-y)^{-p}$. Finally, substituting equation (3.18) into (3.14), yields

$$P(y) = (1-y)^p \left(\sum_{k=0}^{p-1} \begin{pmatrix} p+k-1 \\ k \end{pmatrix} y^k + \mathcal{O}(y) \right) \tag{3.19}$$

where $\mathcal{O}(y) \equiv 0$ in the case of the daublets [20].

To relate now $P(y)$ to $P(\omega)$, the following change of variables is used

$$y = \frac{1-\cos\omega}{2} \quad \text{and} \quad 1-y = \frac{1+\cos\omega}{2} \tag{3.20}$$

which leads to

$$P(\omega) = \left(\frac{1+\cos\omega}{2}\right)^{p} \sum_{k=0}^{p-1} \begin{pmatrix} p+k-1 \\ k \end{pmatrix} \left(\frac{1-\cos\omega}{2}\right)^{k} \tag{3.21}$$

$$= \left(\frac{1+e^{i\omega}}{2}\right)^{p} \left(\frac{1+e^{-i\omega}}{2}\right)^{p} \sum_{k=0}^{p-1} \begin{pmatrix} p+k-1 \\ k \end{pmatrix} \left(\frac{1-e^{i\omega}}{2}\right)^{k} \left(\frac{1-e^{-i\omega}}{2}\right)^{k} \tag{3.22}$$

or equivalently in the terms of the Z-transform,

$$P(z) = \left(\frac{1+z}{2}\right)^{p} \left(\frac{1+z^{-1}}{2}\right)^{p} \sum_{k=0}^{p-1} \begin{pmatrix} p+k-1 \\ k \end{pmatrix} \left(\frac{1-z}{2}\right)^{k} \left(\frac{1-z^{-1}}{2}\right)^{k} \tag{3.23}$$

Finally, the determination of m_0 relies on the spectral factorization[11] of $P(z)$. Indeed, by noting that the zeros $R_p(z)$ come either in real duplets, r_k and r_k^{-1}, or in complex quarduplets, $z_l, z_l^*, z_l^{-1}, z_l^{*-1}$, the following factorization is obtained [20],

$$R_p(z) = 2^{-p+1} \begin{pmatrix} 2p-2 \\ p-1 \end{pmatrix}^{\frac{1}{2}} \prod_k \frac{(z-r_k)}{\sqrt{|r_k|}} \prod_l \frac{z^2 + |z_l^2| - 2z\,\mathrm{Re}\, z_l}{|z_l|} \tag{3.24}$$

The following recipe summarizes the design procedure of m_0 :

---

[11] The *Fejèr-Riesz* lemma guaranties that such a factorization is possible [20].

1. *For a given number of vanishing moments p, determine the zeros of $R_p(z)$ using (3.24).*

2. *Choose one zero out of every pair of real zeros r_k, r_k^{-1} and one conjugated pair out of every quarduplet $z_l, z_l^*, z_l^{-1}, z_l^{*-1}$.*

3. *Compute $\mathcal{R}_p(z)$ and substitute it in equation (3.13).*

4. *Find the scaling filter coefficients c_k and normalize them according to (3.2).* ☐

The result is a polynomial in z of degree $2p-1$, corresponding to an orthonormal basis of wavelets in which the basis wavelet ψ given by $\hat{\psi}(\omega) = e^{-\frac{i\omega}{2}} m_0^*\left(\frac{\omega}{2}+\pi\right)\hat{\phi}\left(\frac{\omega}{2}\right)$ (see section 2.4.3) has support width $2p-1$ and p vanishing moments.

Example 3 *Let us illustrate the recipe for the case of $p = 2$, i.e. two vanishing moments. From equation (3.19), one gets*

$$P(y) = (1-y)^2(1+2y)$$

Using the change of variable[12] $y = (1-(z+z^{-1})/2)/2$, *one can directly get the expression of the polynomial P in the Z-domain as,*

$$P(z) = \left(\frac{1+z}{2}\right)^2 \left(\frac{1+z^{-1}}{2}\right)^2 \frac{1}{2}(-z+4-z^{-1})$$

The roots of $R(z)$ are $r_1 = 2+\sqrt{3}$ and $r_2 = 2-\sqrt{3}$ and the choice of the one inside

---

[12] *known as the* Joukowski *transformation.*

the unit circle yields the minimum phase $m_0(z)$ *with roots* $(-1, 1, 2-\sqrt{3})$. *Therefore*

$$\begin{aligned} m_0(z) &= \left(\frac{1+z}{2}\right)^2 (z-(2-\sqrt{3})) \\ &= \alpha\left(z^{-1}+1\right)^2(1-(2-\sqrt{3})z^{-1}) \\ &= \frac{1}{4\sqrt{2}}\left((1+\sqrt{3})+(3+\sqrt{3})z^{-1}+(3-\sqrt{3})z^{-2}+(1-\sqrt{3})z^{-3}\right) \\ &= c_0+c_1+c_2+c_3 \end{aligned}$$

where α *is used to normalize the coefficients such that* $\sum c_k = \sqrt{2}$. *The above result is identical to the one obtained in example 1.* ☐

In figure 3-1, the *Daubechies* scaling functions and wavelets are plotted for 2, 4 and 8 vanishing moments. Notice that the higher the number of vanishing moments is, the larger the support of the scaling functions is.

Least asymmetric wavelets

The choice of $m_0(z)$ having all roots with absolute value inside the unit circle leads to wavelets with a very marked asymmetry as noticed in figure 3-1. Indeed symmetry or antisymmetry and orthogonality are conflicting requirements for compactly supported wavelets. It can even be shown that the *Haar* wavelet is the only orthogonal wavelet which is symmetric and compactly supported [20],[92]. Nonetheless, it is possible to design compactly supported orthogonal wavelets that are less asymmetric than the daublets, namely the *symmlets.* In order to generate the least asymmetric wavelets, the roots of P

have to be selected such that the total phase of m_0

$$m_0(\omega) = \left(\frac{1+e^{i\omega}}{2}\right)^p \prod_k (e^{i\omega} - r_k) \prod_l (e^{i\omega} - z_l)(e^{i\omega} - z_l^*) \tag{3.25}$$

is as close as possible to a linear phase. This could be done by estimating the contribution to the total phase of m_0 of each real root r_k as,

$$\Theta_k(\omega) = \arctan\left(\frac{1+r_k}{1-r_k}\tan\frac{\omega}{2}\right) \tag{3.26}$$

and the phase of each complex root $z_l = r_l e^{i\theta_l}$ as,

$$\Theta_l(\omega) = \arctan\left(\frac{(1-r_l^2)\sin\omega}{(1+r_l^2)\cos\omega - 2r_l\cos\theta_l}\right) \tag{3.27}$$

Then the zeros to retain from every quadruplet or duplet are selected in such a way that the total phase $\Theta_{tot}(\omega)$ is as close to zero as possible. As it can be seen from figure 3-2, the least asymmetric wavelets and scaling functions with 2, 4 and 8 vanishing moments exhibit less asymmetry than the daublets with the same number of vanishing moments.

Coiflets

Up to this point the extra terms in equation (3.19) have been dropped. This led to the minimum support $[2p-1]$ for a given number p of wavelets vanishing moments. It happens that by considering the higher degree polynomial $\mathcal{O}(y)$, the subsequent extra degree of freedom can be used for the design of orthogonal wavelets whose scaling functions

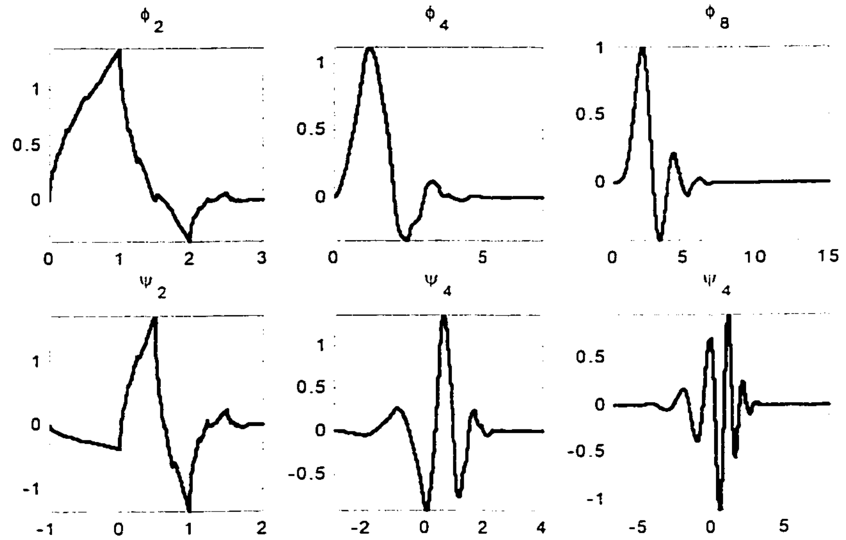


Figure 3-1: Most-asymmetric scaling function ϕ_p and wavelet ψ_p with p vanishing moments.

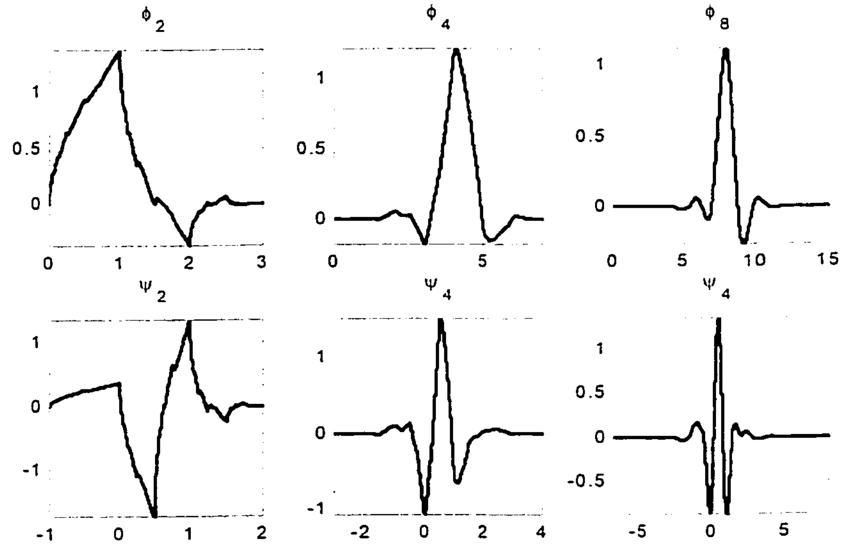


Figure 3-2: Least-asymmetric scaling function ϕ_p and wavelet ψ_p with p vanishing moments.

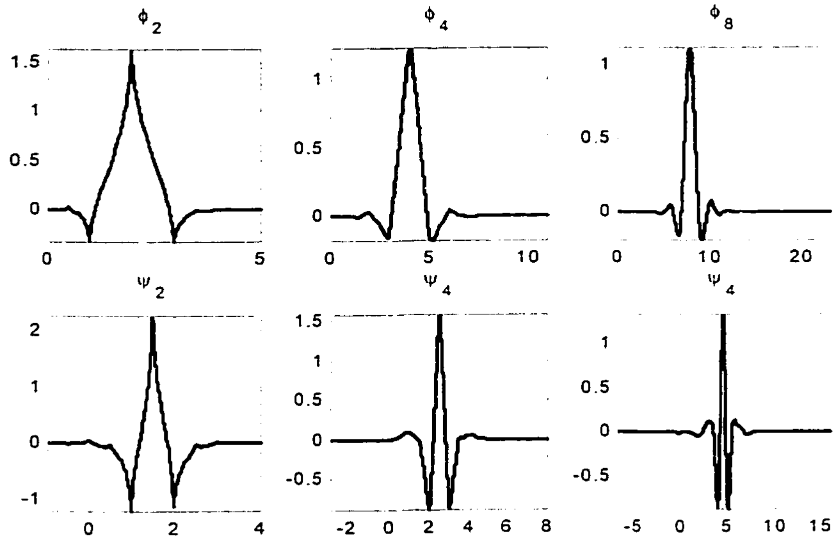


Figure 3-3: Coiflet scaling function ϕ_p and wavelet ψ_p with p vanishing moments.

satisfy more moments conditions than the zero moment condition[13], i.e.$\int \phi(t)dt = 1$. The combination of zero wavelet moments and zero scaling function moments gives a better approximation of the finest scale wavelet coefficients by the samples of the signal $f(t)$, i.e. $\langle \phi_{-J_1k}, f\rangle \approx 2^{\frac{J_1}{2}} f(2^{-J_1}k)$ for large J_1, with an error that is negligibly small where $f(t)$ is smooth.

More precisely the scaling functions and the wavelets are designed so that

$$\int_{-\infty}^{\infty} \phi(t)dt = 1 \tag{3.28}$$

$$\int_{-\infty}^{\infty} t^m\phi(t)dt = 0, \quad m = 1, ..., p-1 \tag{3.29}$$

$$\int_{-\infty}^{\infty} t^m\psi(t)dt = 0, \quad m = 0, ..., p-1 \tag{3.30}$$

---

[13] which ensures that the scaling function has a lowpass characteristic.

Moreover coiflets are smoother and more symmetric than the daublets and symmlets but this comes at the expense of a wider support $[3p-1]$ for the basis functions ϕ_p and ψ_p. In figure 3-3, the scaling functions and wavelets with 2, 4 and 8 vanishing moments are represented. When compared to the daublets of figure 3-1 and to the symmlets of figure 3-2, it is evident that the bases are much more symmetric but the support is wider.

3.3 Biorthogonal Wavelet System

Requiring the wavelet system to be orthogonal across both translations and scales gives a clean, robust, and symmetric implementation using filter banks. It places however strong limitations on the possibilities of the system as the orthogonality constraint uses up a large number of the design degrees of freedom. This prevents using linear phase analysis and synthesis filter bank as well as using different analysis and synthesis filters. The biorthogonal wavelet system uses a non-orthogonal basis and its dual to allow a greater flexibility in designing wavelet systems. The lack of orthogonality can be seen as an additional source of errors as some errors will inevitably occur because of the processing of the wavelet coefficients. However, these errors can be easily masked by using more regular synthesis wavelet filters.

3.3.1 Biorthogonal multiresolution

Biorthogonal wavelets are an important generalization of the orthogonal wavelets. Always within the MRA framework, instead of a single orthonormal basis, a pair of *dual* biorthogonal basis functions is constructed, one for the analysis stage and the other for the synthesis one. Recall that, in the context of orthonormal MRA (see section 2.3), the projection operators onto the approximation and detail subspaces V_j and W_j, have been respectively defined as

$$P_{V_j} f = \sum_{k\in\mathbb{Z}} \underbrace{\overbrace{\langle f, \phi_{jk}\rangle}^{\text{analysis}} \phi_{jk}}_{\text{synthesis}}, \qquad \text{and} \qquad P_{W_j} f = \sum_{k\in\mathbb{Z}} \underbrace{\overbrace{\langle f, \psi_{jk}\rangle}^{\text{analysis}} \psi_{jk}}_{\text{synthesis}} \tag{3.31}$$

where the functions ϕ and ψ perform a double task, i.e. computing the wavelet transform coefficients $c_{jk} = \langle f, \phi_{jk}\rangle$ and $d_{jk} = \langle f, \psi_{jk}\rangle$, and reconstructing the projections $P_{V_j} f$ and $P_{W_j} f$ of the signal $f(t)$ onto V_j and W_j from the wavelet transform coefficients.

The more general instance of biorthogonal multiresolution analysis employs similar projection operators, i.e.

$$P_{V_j} f = \sum_{k\in\mathbb{Z}} \underbrace{\overbrace{\langle f, \phi_{jk}\rangle}^{\text{analysis}} \tilde{\phi}_{jk}}_{\text{synthesis}}, \qquad \text{and} \qquad P_{W_j} f = \sum_{k\in\mathbb{Z}} \underbrace{\overbrace{\langle f, \psi_{jk}\rangle}^{\text{analysis}} \tilde{\psi}_{jk}}_{\text{synthesis}} \tag{3.32}$$

where the pair of functions $\{\phi_{jk}\psi_{jk}\}$ and $\left\{\tilde{\phi}_{jk}, \tilde{\psi}_{jk}\right\}$ are used to share the workload such that the first pair is performing the analysis or decomposition, while the other performs the synthesis or reconstruction of the signal.

The functions ϕ and ψ are called the *primal* scaling function and wavelet, whereas the functions $\tilde{\phi}$ and $\tilde{\psi}$ are called the *dual* scaling function and wavelet[14] respectively. In terms of the MRA, this new decomposition/reconstruction scheme leads to a family of biorthogonal scaling functions and wavelets that are dual bases

---

[14] The role of these functions can however be interchanged by virtue of the *duality principle*.

of the approximation and detail spaces. More precisely, we define a pair of scaling functions ϕ_j and $\tilde{\phi}_j$ that forms a basis of the subpaces V_j and $\tilde{V}_j$ respectively. Similarly, we define a pair of wavelets ψ_j and $\tilde{\psi}_j$ that also forms a basis of the subpaces W_j and $\tilde{W}_j$. These pairs of functions generate dual MRA ladders, such that

$$\{0\} \subset \; ... \subset V_1 \subset V_0 \subset V_{-1} \subset ... \subset L^2(\mathbb{R}) \tag{3.33}$$

$$\{0\} \subset \; ... \subset \tilde{V}_1 \subset \tilde{V}_0 \subset \tilde{V}_{-1} \subset ... \subset L^2(\mathbb{R}) \tag{3.34}$$

where $V_0 = Span\{\phi_{0,k} \backslash k \in \mathbb{Z}\}$ and $\tilde{V}_0 = Span\{\tilde{\phi}_{0,k} \backslash k \in \mathbb{Z}\}$. The spaces W_j and $\tilde{W}_j$ generated by ψ_j and $\tilde{\psi}_j$ are respectively the components of V_j in V_{j-1} and of $\tilde{V}_j$ in $\tilde{V}_{j-1}$. In other words, $V_{j-1} = V_j \oplus W_j$ and $\tilde{V}_{j-1} = \tilde{V}_j \oplus \tilde{W}_j$. The intersection of these spaces is null, i.e. $V_j \cap W_j = \{0\}$ and $\tilde{V}_j \cap \tilde{W}_j = \{0\}$, but the spaces V_j, W_j and $\tilde{V}_j, \tilde{W}_j$ are not orthogonal.

In order to compensate for the lack of orthogonality within the approximation and detail spaces, a biorthogonality relation is imposed between the primal and dual multiresolution ladders, such that

$$V_j \perp \tilde{W}_j \quad \text{and} \quad \tilde{V}_j \perp W_j \tag{3.35}$$

and consequently

$$W_j \perp \tilde{W}_l \quad \text{for} \quad j \neq l \tag{3.36}$$

The two multiresolution hierarchies and their sequences of complement spaces fit together according to an interlaced pattern. As a consequence, the above biorthogonality conditions implies that the bases of the underlying spaces must relate as

$$\begin{aligned} \left\langle \tilde{\phi}(t), \psi(t-k) \right\rangle &= \int \tilde{\phi}(t)\psi^*(t-k)dt = 0 && (3.37) \\ \left\langle \tilde{\psi}(t), \phi(t-k) \right\rangle &= \int \tilde{\psi}(t)\phi^*(t-k)dt = 0 && (3.38) \end{aligned}$$

and

$$\begin{aligned} \left\langle \tilde{\phi}(t), \phi(t-k) \right\rangle &= \int \tilde{\phi}(t)\phi^*(t-k)dt = \delta_k && (3.39) \\ \left\langle \tilde{\psi}(t), \psi(t-k) \right\rangle &= \int \tilde{\psi}(t)\psi^*(t-k)dt = \delta_k && (3.40) \end{aligned}$$

3.3.2 Biorthogonal filters conditions

The two pairs of scaling functions and wavelets, ϕ, ψ and $\tilde{\phi}, \tilde{\psi}$ are defined recursively by the two pairs of polynomial filters m_0, m_1 and $\tilde{m}_0, \tilde{m}_1$ given in the frequency domain[15] by

$$\begin{matrix} \phi(\omega) = m_0\left(\frac{\omega}{2}\right)\phi\left(\frac{\omega}{2}\right) & ; \psi(\omega) = m_1\left(\frac{\omega}{2}\right)\phi\left(\frac{\omega}{2}\right) \\ \tilde{\phi}(\omega) = \tilde{m}_0\left(\frac{\omega}{2}\right)\tilde{\phi}\left(\frac{\omega}{2}\right) & ; \tilde{\psi}(\omega) = \tilde{m}_1\left(\frac{\omega}{2}\right)\tilde{\phi}\left(\frac{\omega}{2}\right) \end{matrix} \qquad (3.41)$$

---

[15] The ˆ of the *Fourier* transform will be omitted.

where

$$
\begin{aligned}
m_0(\omega) &\equiv \tfrac{1}{\sqrt{2}} \textstyle\sum_{k\in\mathbb{Z}} c_k e^{i\omega k} \quad ; m_1(\omega) \equiv \tfrac{1}{\sqrt{2}} \textstyle\sum_{k\in\mathbb{Z}} b_k e^{i\omega k} \\
\tilde{m}_0(\omega) &\equiv \tfrac{1}{\sqrt{2}} \textstyle\sum_{k\in\mathbb{Z}} \tilde{c}_k e^{i\omega k} \quad ; \tilde{m}_1(\omega) \equiv \tfrac{1}{\sqrt{2}} \textstyle\sum_{k\in\mathbb{Z}} \tilde{b}_k e^{i\omega k}
\end{aligned}
\tag{3.42}
$$

By computing the *Fourier* transform of the inner products in equations (3.37)-(3.40) and using the same arguments as in section 3.2.1, the biorthogonality conditions are expressed in the frequency domain as

$$\sum_k \tilde{\phi}(\omega + 2k\pi)\phi^*(\omega + 2k\pi) = 1 \tag{3.43}$$

$$\sum_k \tilde{\psi}(\omega + 2k\pi)\psi^*(\omega + 2k\pi) = 1 \tag{3.44}$$

$$\sum_k \tilde{\psi}(\omega + 2k\pi)\phi^*(\omega + 2k\pi) = 0 \tag{3.45}$$

$$\sum_k \tilde{\phi}(\omega + 2k\pi)\psi^*(\omega + 2k\pi) = 0 \tag{3.46}$$

This means that the polynomial filters and their dual have to satisfy the following conditions,

$$\tilde{m}_0(\omega)m_0(\omega) + \tilde{m}_0(\omega + \pi)m_0(\omega + \pi) = 1 \tag{3.47}$$

$$\tilde{m}_1(\omega)m_1(\omega) + \tilde{m}_1(\omega + \pi)m_1(\omega + \pi) = 1 \tag{3.48}$$

$$\tilde{m}_1(\omega)m_0(\omega) + \tilde{m}_1(\omega + \pi)m_0(\omega + \pi) = 0 \tag{3.49}$$

$$\tilde{m}_0(\omega)m_1(\omega) + \tilde{m}_0(\omega + \pi)m_1(\omega + \pi) = 0 \tag{3.50}$$

3.3.3 Design of biorthogonal wavelet filters

To design the polynomial filters $m_0, m_1, \tilde{m}_0, \tilde{m}_1$ that generate the biorthogonal wavelets, the same procedure as in section 3.2.2 is used, except that the reciprocal product polynomial is now defined as

$$P(\omega) = \tilde{m}_0(\omega) m_0(\omega) \tag{3.51}$$

Note that the analysis filters can be different from the synthesis filters, because there are two biorthogonal MRA hierarchies. This offers more freedom for the design of linear phase FIR scaling and wavelets filters, i.e. symmetric wavelets. This gain in flexibility stems from the factorization of the polynomial P into two different polynomials, i.e. m_0 and $\tilde{m}_0$, instead of the square root of a single polynomial m_0. Moreover, in order to ensure the regularity of the underlying scaling functions, the polynomials m_0 and $\tilde{m}_0$ must have at least one zero at the *Nyquist* frequency π, i.e.

$$P(\omega) = \left(\left(\frac{1+e^{i\omega}}{2}\right)^p\right)^2 \mathcal{R}_p(\omega) \tag{3.52}$$

The polynomials m_1 and $\tilde{m}_1$ follow from the relationship (2.64) [20],

$$\begin{aligned} m_1(\omega) &= e^{-i\omega} \tilde{m}_0^*(\omega + \pi) && (3.53) \\ \tilde{m}_1(\omega) &= e^{-i\omega} m_0^*(\omega + \pi) && (3.54) \end{aligned}$$

Depending on the factorization of the polynomial $P(\omega)$, the following two subfamilies of biorthogonal wavelets, referred to as the CFD wavelets [17] can be constructed.

B-splines wavelets

The polynomial $P(z)$ is factored into the following two polynomials

$$\tilde{m}_0(z) = \left(\frac{1+z}{2}\right)^{\tilde{p}}; \quad \text{with } \tilde{p} = 2\tilde{l} \text{ or } \tilde{p} = 2\tilde{l}+1 \tag{3.55}$$

and

$$m_0(z) = \left(\frac{1+z}{2}\right)^{p} \sum_{q=0}^{l+\tilde{l}+\sigma} \begin{pmatrix} l+\tilde{l}+\sigma+q \\ q \end{pmatrix} \left(\frac{1-z}{2}\right)^{q} \left(\frac{1-z^{-1}}{2}\right)^{q} \tag{3.56}$$

where $p = 2l$ or $p = 2l+1$, and $\sigma = 0$ if $\tilde{p}$ is even, and $\sigma = 1$ if $\tilde{p}$ is odd.

The explicit expressions for $\tilde{m}_0$ and m_0 yields a family of *B-splines wavelets*, as the synthesis scaling function $\tilde{\phi}$ itself is a B-spline function[16] of order $\tilde{p}$. Note that $\tilde{m}_0$ depends only on $\tilde{p}$, while m_0 depends on both p and $\tilde{p}$.

The B-splines wavelets family has two main advantages. Firstly, there is a closed form expression for the B-spline scaling functions and wavelets. Secondly, the coefficients for all filters are dyadic rationals, which makes them suitable for fast and exact hardware implementation. Some of them are displayed in figures 3-4 and 3-5 together with the corresponding B-spline wavelets and scaling functions with two dual vanishing moments and two and four primal vanishing moments respectively. However, the major disadvantage of B-splines wavelets is that the analysis and the synthesis (reconstruction) filters have

---

[16] B-splines of order p are piecewise polynomials that come from the convolution of p box functions [24].

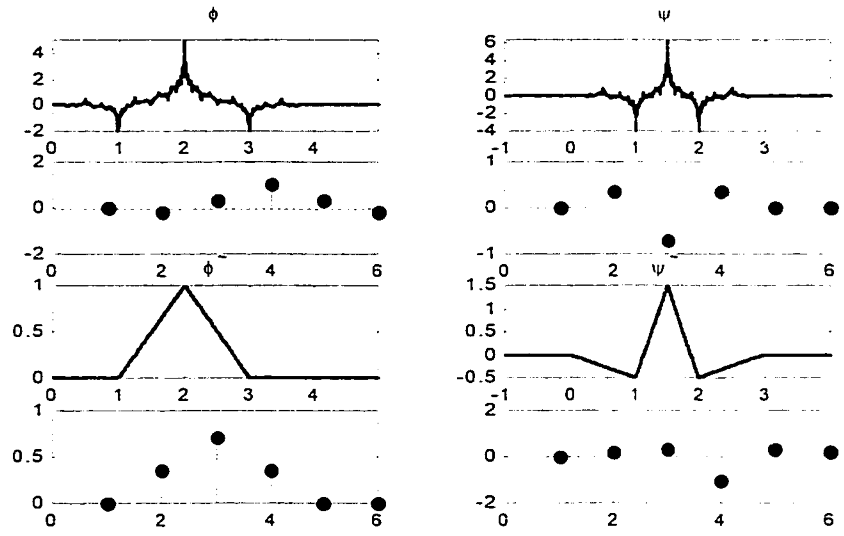


Figure 3-4: B-spline biorthogonal wavelet system and filters coefficients for $\tilde{p} = p = 2$.

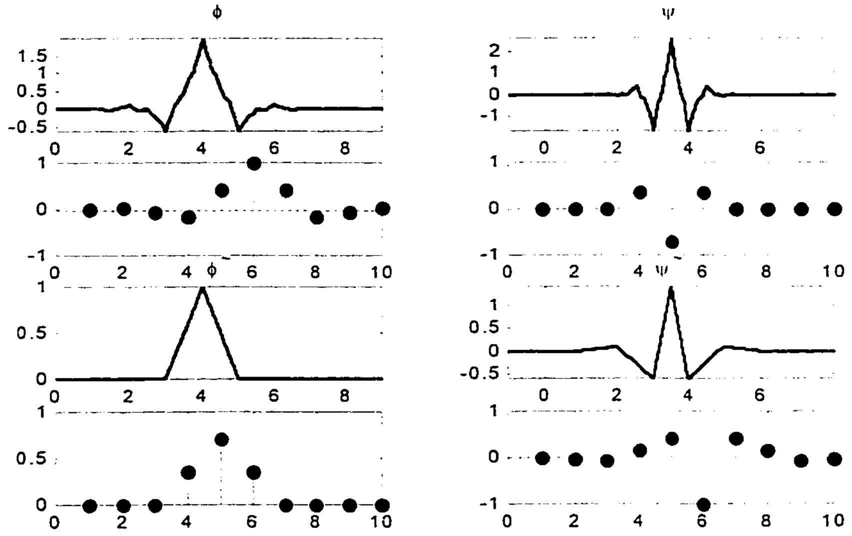


Figure 3-5: B-spline biorthogonal wavelet system and filters coefficients for $\tilde{p} = 2$ and $p = 4$.

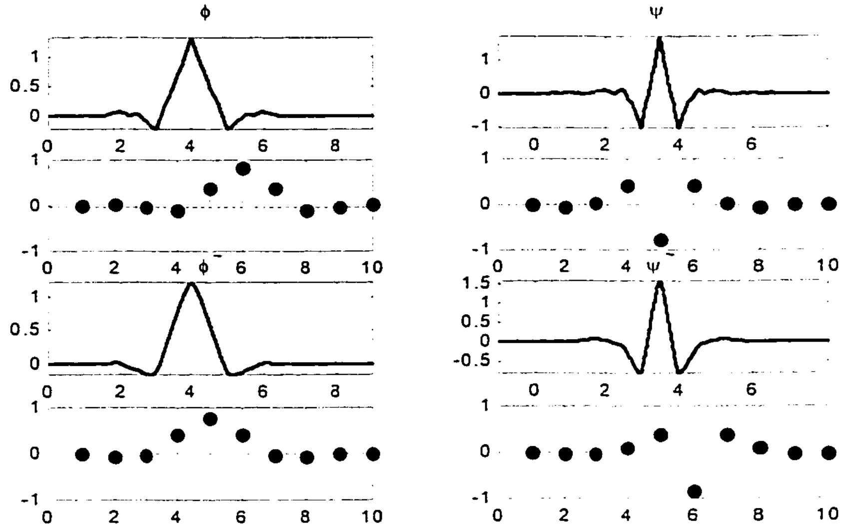


Figure 3-6: Biorthogonal (9-7) taps wavelet system.

very different support widths as it appears in figures 3-4 and 3-5.

Wavelets with closer support widths

It is possible to construct biorthogonal wavelets with less disparate filter lengths by choosing a factorization of the polynomial P, such that the filter functions m_0 and $\tilde{m}_0$ have both linear phase and similar degree. For a fixed $l + \tilde{l}$, all roots of $P(z)$, real zeros r_i and pairs of complex conjugate zeros $\{z_j, z_j^*\}$ are determined such that one can factorize the latter as

$$P(z) = \prod_i (z - r_i) \prod_j (z - z_j)(z - z_j^*) \tag{3.57}$$

Then all the combinations of roots are generated such that $r_i, \{z_j, z_j^*\}$ are assigned either to m_0 or to $\tilde{m}_0$, and select the option $l + \tilde{l}$ that makes the length of m_0 and $\tilde{m}_0$ as close as possible. For $l + \tilde{l} = 4$, the factorization is unique and the corresponding scaling and wavelet filters have 9 and 7 taps respectively as depicted by figure 3-6. These filters are referred to as the FBI wavelet filter because they have been used to establish the FBI wavelet/scalar quantization standard for gray-scale fingerprint image compression [9].

3.4 Two-Dimensional Wavelets

3.4.1 2-D multiresolution

The simplest way to construct wavelet bases for the signal space $L^2(\mathbb{R}^2)$ is to take the product of 1-D wavelet bases as,

$$\Psi_{j_1 j_2, k_1 k_2}(x, y) = \psi_{j_1 k_1}(x)\psi_{j_2 k_2}(y) \tag{3.58}$$

This approach did not retain the attributes of the MRA because the wavelet functions $\psi_{j_1 k_1}\psi_{j_2 k_2}$ mix information at the two different scales 2^{j_1} and 2^{j_2}. To avoid this intricate pattern, we define a separable multiresolution that leads to the construction of separable wavelet bases whose elements are products of wavelets that are dilated at the unique scale j. The first step consists in defining the 2-D subspace $\mathbf{V}_j$ as the tensor product

$$\mathbf{V}_j = V_j \otimes V_j = \overline{span\{f(x)g(x), f, g \in V_j\}} \tag{3.59}$$

In general, the tensor product is used to extend spaces of 1-D signals to spaces of multidimensional signals $f(t)$, or in the case of 2-D signals $f(x, y)$ where x and y define the $x-y$ plane. The tensor product satisfies linearity and distributivity properties. Moreover, in the case of *Hilbert spaces* $\mathbb{H}$, an inner product in $L^2(\mathbb{R}^{\mathcal{K}}) \subset \mathbb{H}$ is derived from inner products in $L_1^2(\mathbb{R})$ and $L_2^2(\mathbb{R})$ as

$$\langle \alpha_1 \otimes \alpha_2, \beta_1 \otimes \beta_2 \rangle_{L^2(\mathbb{R}^{\mathcal{K}})} = \langle \alpha_1 \otimes \beta_1 \rangle_{L_1^2(\mathbb{R})} \langle \alpha_2 \otimes \beta_2 \rangle_{L_2^2(\mathbb{R})} \tag{3.60}$$

resulting thus, in a straightforward extension of a 1-D MRA, where the same monotonicity, completeness, shift-invariance and scaling properties automatically hold. Consequently, the separable scaling basis Φ defined by

$$\begin{aligned} \Phi_{j,k_1k_2}(x, y) &= \phi_{jk_1}(x)\phi_{jk_2}(y) \\ &= 2^{-j}\phi(2^{-j}x - k_1)\phi(2^{-j}y - k_2) \end{aligned} \tag{3.61}$$

yields a basis for $\mathbf{V}_j$ provided that the scaling function $\phi(.)$ is so for V_j. This basis is obtained by scaling the function $\phi(x)$ and $\phi(y)$ by a factor 2^j and then translating them on a 2-D square dyadic grid, i.e. with intervals 2^j.

To complete the above extension, we characterize the orthogonal complement spaces that catch the details lost when going from a scale to a coarser one as in the 1-D case, by defining $\mathbf{W}_j$ to be the orthogonal complement of $\mathbf{V}_j$ in $\mathbf{V}_{j-1}$ as

$$\begin{aligned}
\mathbf{V}_{j-1} &= V_{j-1} \otimes V_{j-1} \\
&= (V_j \oplus W_j) \otimes (V_j \oplus W_j) \\
&= (V_j \otimes V_j) \oplus \{(V_j \otimes W_j) \oplus (W_j \otimes V_j) \oplus (W_j \otimes W_j) \\
&= \mathbf{V}_j \oplus \mathbf{W}_j \qquad (3.62)
\end{aligned}$$

where we used the distributivity of $\oplus$ with respect to $\otimes$.

From the spaces expansion (3.62), it can be seen that the detail space $\mathbf{W}_j$ is composed of three orthogonal subspaces whose bases are the corresponding tensor products of their components, i.e.

$$\begin{aligned}
\Psi^H_{j,k_1k_2}(x,y) &= \phi_{jk_1}(x)\psi_{jk_2}(y) \\
&= 2^{-j}\phi(2^{-j}x - k_1)\psi(2^{-j}y - k_2) \qquad (3.63) \\
\Psi^V_{j,k_1k_2}(x,y) &= \psi_{jk_1}(x)\phi_{jk_2}(y) \\
&= 2^{-j}\psi(2^{-j}x - k_1)\phi(2^{-j}y - k_2) \qquad (3.64) \\
\Psi^D_{j,k_1k_2}(x,y) &= \psi_{jk_1}(x)\psi_{jk_2}(y) \\
&= 2^{-j}\psi(2^{-j}x - k_1)\psi(2^{-j}y - k_2) \qquad (3.65)
\end{aligned}$$

To give an interpretation of the above *anisotropic* 2-D wavelets and make the meaning of the superscripts V, H and D relevant, we consider the case of digital images. These three wavelets extract image details at different scales and orientations. Indeed,

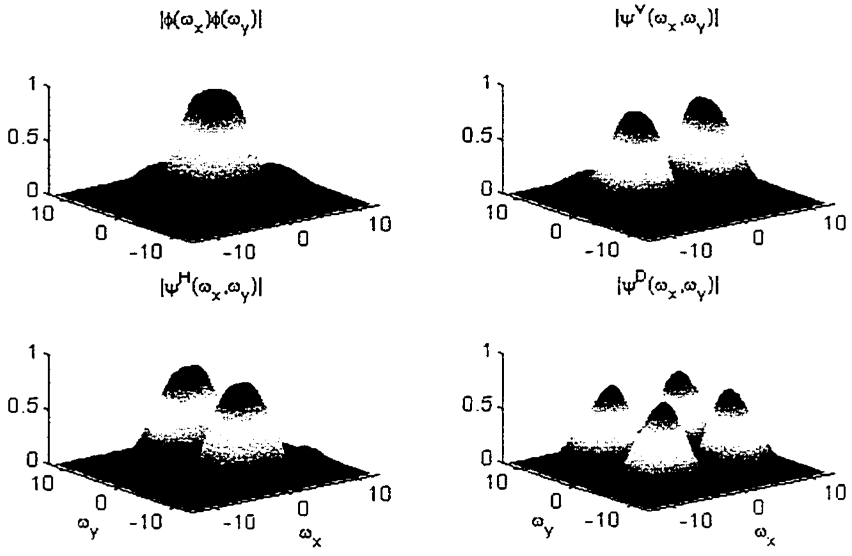


Figure 3-7: Fourier transforms of the separable 2-D daublet-2 wavelet system.

because of their lowpass and highpass nature over positive frequencies, $\hat{\phi}(\omega)$ and $\hat{\psi}(\omega)$ have their energy concentrated respectively in the interval $[0,\pi]$ and $[\pi,2\pi]$. Therefore, $\left|\hat{\Psi}^H(\omega_x,\omega_y)\right|$ is large at low horizontal frequencies ω_x and high vertical frequencies ω_y, whereas $\left|\hat{\Psi}^V(\omega_x,\omega_y)\right|$ is large at high horizontal frequencies ω_x and low vertical frequencies ω_y. Finally $\left|\hat{\Psi}^D(\omega_x,\omega_y)\right|$ is large at high horizontal and vertical frequencies [68]. These facts are illustrated by figure 3-7 with the *Fourier* transforms of the separable 2-D scaling and the three anisotropic *daublet-2* wavelets of figure 3-8.

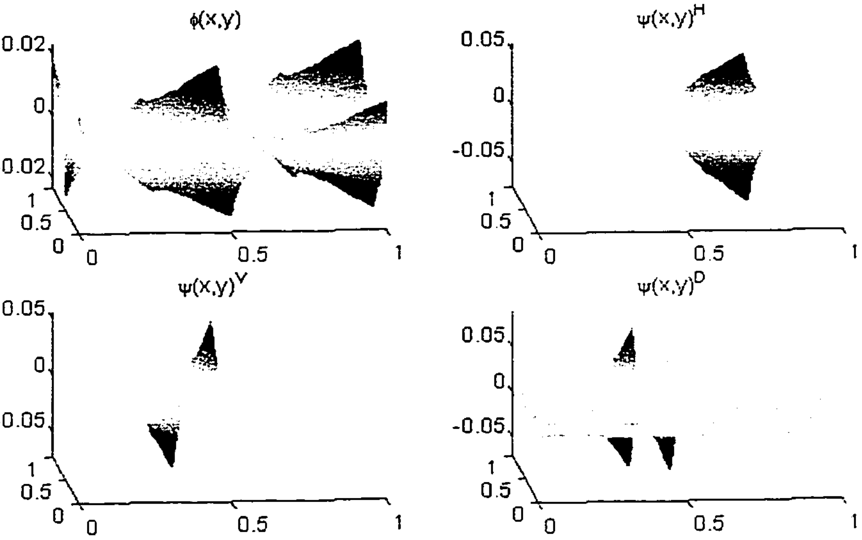


Figure 3-8: 2-D separable daublet-2 basis functions.

3.4.2 2-D wavelet approximation

The 2-D wavelet approximation is a straightforward generalization of the 1-D case where a 2-D signal $f(x, y)$ can be written as

$$\begin{aligned} f(x,y) \approx & \sum_{k_1,k_2 \in \mathbb{Z}} a_{J_0 k_1 k_2} \Phi_{j,k_1 k_2}(x,y) + \sum_{j=J_0}^{J_1+1} \sum_{k_1,k_2 \in \mathbb{Z}} d^V_{jk_1k_2} \Psi^V_{j,k_1k_2}(x,y) \\ & + \sum_{j=J_0}^{J_1+1} \sum_{k_1,k_2 \in \mathbb{Z}} d^H_{jk_1k_2} \Psi^H_{j,k_1k_2}(x,y) + \sum_{j=J_0}^{J_1+1} \sum_{k_1,k_2 \in \mathbb{Z}} d^D_{jk_1k_2} \Psi^D_{j,k_1k_2}(x,y) \quad (3.66) \end{aligned}$$

where the signal $f(x, y)$ is decomposed into a sum of coarse resolution (at the scale J_0) smooth coefficients and a sum of fine to coarse resolution (scales J_0 to $J_1 + 1$) detail coefficients. However in the 2-D case there are three types of detail coefficients associated with the three wavelet bases, namely the vertical, the horizontal, and the diagonal detail coefficients. Moreover, the following 2-D functions

$$\begin{aligned} A_{J_0}(x,y) &= \sum_{k_1,k_2\in\mathbb{Z}} a_{J_0k_1k_2}\Phi_{j,k_1k_2}(x,y) && (3.67)\\ D_j^V(x,y) &= \sum_{k_1,k_2\in\mathbb{Z}} d^V_{jk_1k_2}\Psi^V_{j,k_1k_2}(x,y) && (3.68)\\ D_j^H(x,y) &= \sum_{k_1,k_2\in\mathbb{Z}} d^H_{jk_1k_2}\Psi^H_{j,k_1k_2}(x,y) && (3.69)\\ D_j^D(x,y) &= \sum_{k_1,k_2\in\mathbb{Z}} d^D_{jk_1k_2}\Psi^D_{j,k_1k_2}(x,y) && (3.70) \end{aligned}$$

are called respectively the residual image, the vertical details image, the horizontal image, and the diagonal image. The wavelet series approximation (3.66) becomes thus,

$$f(x,y) \approx A_{J_0}(x,y) + \sum_{j=J_0}^{J_1+1} D_j^V(x,y) + \sum_{j=J_0}^{J_1+1} D_j^H(x,y) + \sum_{j=J_0}^{J_1+1} D_j^D(x,y) \qquad (3.71)$$

Finally, the 2-D wavelet transform coefficients are given approximately by the integrals

$$\begin{aligned} a_{J_0k_1k_2} &\approx \int\int \Phi_{j,k_1k_2}(x,y)f(x,y)dxdy && (3.72)\\ d^V_{jk_1k_2} &\approx \int\int \Psi^V_{j,k_1k_2}(x,y)f(x,y)dxdy && (3.73)\\ d^H_{jk_1k_2} &\approx \int\int \Psi^H_{j,k_1k_2}(x,y)f(x,y)dxdy && (3.74)\\ d^D_{jk_1k_2} &\approx \int\int \Psi^D_{j,k_1k_2}(x,y)f(x,y)dxdy && (3.75) \end{aligned}$$

but in practice, these are computed using the *Mallat* algorithm, which is implemented with a 2-D two-channel filter bank. This issue will be covered in the next chapter.

Chapter 4

FAST WAVELET TRANSFORM

The present chapter bridges the gap between the theory and the implementation of the DWT. Indeed, computational complexity is what makes the difference between an impractical and a useful algorithm like the *fast wavelet transform* (FWT). The DWT is implemented efficiently by exploiting the scaling property of the multiresolution analysis theory where the ensuing recursion does only constitute the machinery behind the calculation of the DWT but it also unravels the connections between the multiresolution analysis and filter banks [100]. First the *Mallat* algorithm is discussed before establishing the connections between the DWT and the *two-channel filter bank*. Then the different approaches to extend the 1-D DWT to the 2-D case are presented. In general, it is common to use a separable 2-D DWT where filter banks are applied to the each dimension individually. This can be achieved along two routes. The first one consists of first applying the 1-D DWT to each row of the input data matrix, e.g. pixels in the case of digital images. Next, the transformed rows are treated as if they were themselves a new image and the DWT is applied again but to each column. This leads to the so-called *standard wavelet decomposition.* The second type of separable 2-D DWT is called the *non-standard decomposition.* It alternates between operations on rows and columns, by first performing one step of pairwise 1-D DWT on the pixel values of each row of the image matrix. Next, a similar pairwise 1-D DWT is applied to each column of the result. To complete the transformation, the same process is repeated recursively only on the quadrant containing averages in both directions. As the non-standard wavelet expansion is computational efficient and simple [91], it is exclusively used in the present work. The last section of this chapter concerns borders corrections. In theory the DWT operator maps an infinite dimension signal $f(t)$ to an infinite set of wavelet coefficients, but seismic data do not extend indefinitely in time or space but are finite length signals. Consequently, the DWT operator may be truncated to a finite and invertible one or alternatively, the finite signal may be extended to an infinite one. In both cases, the resulting transform differs at the boundaries, while remaining the same for data far enough from the boundaries. Moreover, truncating the wavelet operator means designing the so-called *wavelets on the interval* [18], which have proven to be effective in theory but are not satisfactory from a practical point of view [25],[68]. It is often preferable to use simple schemes based on signal extension at the boundaries. These methods include: *zero-padding*, *periodic* and *symmetric extensions.*

4.1 Discrete Multiresolution Theory

Discrete multiresolution theory is considered as the unified framework including *discrete wavelet transforms*, *two-channel filter banks* and *pyramid transforms.* Additionally, it is well suited for digital signal processing and does not require the use of functional spaces as an intermediate step. Indeed, unlike the continuous-time case, where the scaling function $\phi(t)$ and the wavelet $\psi(t)$ are brought into play, the discrete MRA involves only discrete scaling and wavelet filters designed as per the techniques discussed in the previous chapter. The notions of dilation and time-shift take thus another significance where the concepts of scale and resolution are revisited in the context of multirate signal processing theory. We begin by reviewing the pioneering pyramid multiresolution schemes, then the *Mallat* pyramid algorithm that is behind the FWT is discussed. The filter bank theory is then reviewed emphasizing the iterated two-channel filter bank that naturally yields a discrete-time implementation of the FWT.

4.1.1 Pioneering multiresolution schemes

Two methods that lead to the DWT were developed independently in the early eighties. These are the *pyramid* [13] and the *subband* [39] multiresolution techniques proposed for the coding of images and speech respectively.

Pyramid multiresolution

Given the original sequence $a_n^{j-1}, n \in \mathbb{Z}$, a lower resolution signal is obtained by lowpass filtering with a *halfband* lowpass filter having an impulse response h_n. Following *Nyquist*'s rule, we can *subsample* by two, i.e. dropping every other sample, thus doubling the scale in the analysis. This results in a sequence a^j given by

$$a_n^j = \sum_{k \in \mathbb{Z}} a_{2n-k}^{j-1} h_k \tag{4.1}$$

where the resolution change occurs because of the lowpass filtering, i.e. loss of high frequency details, whereas the scale change is due to the subsampling (or decimation) by two since a shift by two in the original signal a_n^{j-1} results in a shift by one in the new signal a_n^j.

The multiresolution pyramid algorithm consists in finding an approximation $\breve{a}_n^{j-1}$ to the original signal based on the lowpass and subsampled version. This is done by first bringing back a_n^j to the original scale j by upsampling by two, i.e. $\breve{a}_{2n}^{j\prime} = a_n^j$. Then $\breve{a}_n^{j\prime}$ is interpolated with a filter with an impulse response g_n' to obtain the approximation

$$\breve{a}_n^{j-1} = \sum_{k \in \mathbb{Z}} a_{n-k}^{j\prime} g_k' \tag{4.2}$$

In general $\breve{a}_n^{j-1} \neq a_n^{j-1}$ and the difference is given by

$$d_n^{j-1} = a_n^{j-1} - \breve{a}_n^{j-1} \tag{4.3}$$

This scheme can be iterated on $a_n^j, a_n^{j+1}, \ldots$, creating thus a hierarchy of differences and lower resolution signals where the resulting signals are getting shorter and shorter. Whence the denomination pyramid multiresolution.

Note that the pyramid scheme has some redundancy since at one stage of the decomposition a signal with sampling rate f_s is mapped into two signals d_n^j and a_n^{j-1} with sampling rates f_s and $f_s/2$ respectively, resulting in an increase of 50% in the number of samples at the first stage, plus 25% at the second stage and so on. Elsewhere, in the case of ideal halfband lowpass filter, it turns out that the signal d_n^{j-1} contains exactly the frequencies of a_n^j that are above $\pi/2$. Consequently, it can be subsampled by two as well without any loss of information. This hints at the fact that critically sampled scheme may exist and takes us to the second scheme.

Subband multiresolution

The subband multiresolution is conceptually similar to the pyramid scheme as the original signal is split into a coarse approximation and a detail signal. The lowpass subsampled approximation is obtained exactly as in equation (4.1), but instead of the full bandwidth difference signal d_n^{j-1}, the added detail are computed as a highpass filtered version of a_n^{j-1}, using a filter with an impulse response g_n such that

$$d_n^j = \sum_{k \in \mathbb{Z}} a_{2n-k}^{j-1} g_k \tag{4.4}$$

It is clear that if h_n is an ideal halfband lowpass filter, then an ideal halfband bandpass filter will lead to a perfect decomposition of the original signal into the two subsampled versions, i.e. a_n^j and d_n^j. This is essentially one stage of the DWT using scaling and wavelet filters, since the input signal is mapped into a lowpass resolution approximation (at twice the scale) and an added detail signal (also at twice the scale) containing the detail left out when moving from a scale $j-1$ to a higher (coarser) scale j. However, the reconstructed signal

$$\breve{a}_n^{j-1} = \sum_{k\in\mathbb{Z}} a_{2n-k}^{j-1} h_k + \sum_{k\in\mathbb{Z}} a_{2n-k}^{j-1} g_k \tag{4.5}$$

is not identical to the original signal a^{j-1} unless the filters h_n and g_n meet some specific constraints. Such a subband multiresolution is best studied within the framework of the perfect reconstruction two-channel filter banks introduced later.

4.1.2 Fast wavelet transform

The fast wavelet transform is based on the *Mallat* algorithm that finds its roots in computer vision [81]. The forward transform converts the samples of a signal $f(t)$ into a set of wavelet coefficients. Conversely, the inverse transform reconstructs the signal from its wavelet representation back to the time (spatial) domain. The FWT is executed in linear time $\mathcal{O}(N)$, which is faster than the FFT that is performed in logarithmic time $N\log(N)$. The decomposition and reconstruction *Mallat* algorithms are discussed next.

Decomposition Mallat algorithm

Recall that the wavelet decomposition of the signal $f(t)$ is performed according to equation (2.45) where the orthogonal projections on the scale and wavelet (sub-) spaces are given by relations (2.29) and (2.38) respectively. These relations rely however, on the explicit calculation of the inner products using integrals (2.30) and (2.39). It was hinted there that the recursive wavelet decomposition/reconstruction scheme involves only projections between two consecutive subspaces, say $V_j \subset V_{j-1}$ and $W_j \subset V_{j-1}$ of the MRA ladder.

Mathematically, these relationships are given by the two-scale difference equations (2.52) and (2.54), which are reproduced here for convenience,

$$\phi(t) = \sum_{m\in\mathbb{Z}} c_m \sqrt{2}\phi(2t - m) \tag{4.6}$$

$$\psi(t) = \sum_{m\in\mathbb{Z}} b_m \sqrt{2}\phi(2t - m) \tag{4.7}$$

Using these two relations, we can express the basis functions of the scale and wavelet spaces V_j and W_j at scale j, that are defined by

$$\phi_k^j(t) = 2^{-\frac{j}{2}}\phi(2^{-j}t - k) \tag{4.8}$$

$$\psi_k^j(t) = 2^{-\frac{j}{2}}\psi(2^{-j}t - k) \tag{4.9}$$

in terms of the basis functions of the subsequent scale space V_{j-1} at finer scale $j-1$. For, substituting equation (4.6) into equation (4.8) yields

$$\begin{aligned} \phi_k^j(t) &= 2^{-\frac{j}{2}} \sum_{m\in\mathbb{Z}} c_m \sqrt{2}\phi(22^{-j}t - 2k - m) \\ &= \sum_{m\in\mathbb{Z}} c_m 2^{-\frac{j}{2}}\sqrt{2}\phi(2^{-j+1}t - m - 2k) \\ &= \sum_{m\in\mathbb{Z}} c_m 2^{\frac{-j+1}{2}}\phi(2^{-j+1}t - m - 2k) \end{aligned} \tag{4.10}$$

or

$$\phi_k^j(t) = \sum_{m\in\mathbb{Z}} c_m \phi_{m+2k}^{j-1}(t) \tag{4.11}$$

which by making the change of variable $n = 2k + m$, becomes

$$\phi_k^j(t) = \sum_{n\in\mathbb{Z}} c_{n-2k} \phi_n^{j-1}(t) \tag{4.12}$$

Finally, by taking the inner product of the signal $f(t)$ with both sides of relation (4.12),

$$\left\langle f, \phi_k^j \right\rangle = \sum_{n\in\mathbb{Z}} c_{n-2k}^* \left\langle f, \phi_n^{j-1} \right\rangle$$

the following recursive equation is obtained

$$a_k^j = \sum_{n\in\mathbb{Z}} c_{n-2k}^* a_k^{j-1} \equiv H_0 a_k^{j-1} \tag{4.13}$$

It expresses the approximation coefficients at a given scale j in terms of the ones at the next finer scale $j-1$.

Similarly, substituting equation (4.7) into equation (4.9) yields

$$\psi_k^j(t) = \sum_{n\in\mathbb{Z}} b_{n-2k} \phi_n^{j-1}(t) \tag{4.14}$$

which leads to the recursive relation relating the details coefficients at scale j and the approximation coefficients at the next finer scale $j-1$, i.e.

$$\begin{tikzcd}
a^{J_1} \arrow[r, "H_0"] \arrow[dr, "H_1"] & a^{J_1+1} \arrow[r, "H_0"] \arrow[dr, "H_1"] & \cdots \arrow[r, "H_0"] \arrow[dr, "H_1"] & a^{J_0-1} \arrow[r, "H_0"] \arrow[dr, "H_1"] & a^{J_0} \\
 & d^{J_1+1} & \cdots & d^{J_0-1} & d^{J_0}
\end{tikzcd}$$

Figure 4-1: Decomposition *Mallat* algorithm.

$$d_k^j = \sum_{n\in\mathbb{Z}} b^*_{n-2k} a_k^{j-1} \equiv H_1 a_k^{j-1} \tag{4.15}$$

Note that H_0 and H_1 stand for the lowpass and the highpass discrete filters respectively.

Equations (4.13) and (4.15) represent the *decomposition (or analysis) Mallat* algorithm [68], which is behind the direct FWT. The decomposition algorithm starts with $l = 2^{J_1}$ approximation wavelet coefficients $\left\{a_k^{J_1}\right\}_l$ at the finest scale J_1. These are in general[1] fairly approximated by the samples of the signal $2^{\frac{J_1}{2}} f(2^{-J_1}k)$. Then the *Mallat* algorithm generates a sequence of detail wavelet coefficients at coarser scales, i.e. $\left\{d_k^{J_1+1}\right\}_{l/2}, \left\{d_k^{J_1+2}\right\}_{l/4}, \ldots, \left\{d_k^{J_0}\right\}_{l/2^{J_0}}$ and the approximation wavelet coefficients at the coarsest scale J_0, i.e. $\left\{a_k^{J_0}\right\}_{l/2^{J_0}}$ as illustrated by figure 4-1. Note that because of the factor $2k$ in the discrete convolutions (4.13) and (4.15), only the even coefficients are retained at each step, yielding the same number of coefficients, i.e. $\frac{l}{2} + \frac{l}{4} + \ldots + \frac{l}{2^{J_0}} + \frac{l}{2^{J_0}} = l$ as the initial sequence.

---

[1] Other choices are discussed in [85].

Reconstruction Mallat algorithm

The reconstruction part of the *Mallat* algorithm generates the coefficients of the scale representation a^{j-1} from the coefficients of the wavelet representation a^j and d^j. Recall that

$$\begin{aligned} f^{j-1} &\equiv P_{V_{j-1}} f \\ &= P_{V_j} f + P_{W_j} f \\ &= \sum_{k\in\mathbb{Z}} a_k^j \phi_k^j + \sum_{k\in\mathbb{Z}} d_k^j \psi_k^j \end{aligned} \tag{4.16}$$

where

$$a_n^{j-1} = \left\langle f^{j-1}, \phi_n^{j-1} \right\rangle \tag{4.17}$$

By inserting equation (4.16) into equation (4.17), one obtains

$$\begin{aligned} a_n^{j-1} &= \left\langle \left(\sum_{k\in\mathbb{Z}} a_k^j \phi_k^j + \sum_{k\in\mathbb{Z}} d_k^j \psi_k^j \right), \phi_n^{j-1} \right\rangle \\ &= \sum_{k\in\mathbb{Z}} a_k^j \left\langle \phi_k^j, \phi_n^{j-1} \right\rangle + \sum_{k\in\mathbb{Z}} d_k^j \left\langle \psi_k^j, \phi_n^{j-1} \right\rangle \end{aligned} \tag{4.18}$$

Using equations (4.12) and (4.14), equation (4.18) can be rewritten as,

$$\begin{aligned} a_n^{j-1} &= \sum_{k\in\mathbb{Z}} c_{n-2k} a_k^j + \sum_{k\in\mathbb{Z}} b_{n-2k} d_k^j \\ &\equiv G_0 a_k^j + G_1 d_k^j \end{aligned} \tag{4.19}$$

$$\begin{tikzcd}
a^{J_0} \arrow[r, "G_0"] & a^{J_0-1} \arrow[r, "G_0"] & \cdots \arrow[r, "G_0"] & a^{J_1+1} \arrow[r, "G_0"] & a^{J_1} \\
d^{J_0} \arrow[ur, "G_1"] & d^{J_0-1} \arrow[ur, "G_1"] & \cdots \arrow[ur, "G_1"] & d^{J_1+1} \arrow[ur, "G_1"] &
\end{tikzcd}$$

Figure 4-2: Reconstruction *Mallat* algorithm.

The equation (4.19) provides, thus a reconstruction formula for the coefficients a_n^{j-1} from the coefficients sequence of the decomposition at level j, for $j = J_0, J_0 - 1, \ldots, J_1 + 1, J_1$ as shown in figure 4-2.

4.2 Filter Banks

The success of the wavelet transform is incontestably due to the multirate *filter banks*. However, we only consider here the two-channel filter bank that is a special case of the more general M-channel filter banks [76],[100].

4.2.1 Two-channel filter bank

A generic 1-D two-channel filter bank is shown in figure 4-3. The input sequence $x(m)$ is first passed through the lowpass filter $H_0(\omega)$ and the highpass filter $H_1(\omega)$. These filters are called the *analysis* filters. As a result, two filtered version of the input signal, i.e. $y_0(m)$ and $y_1(m)$, are generated. Due to bandpass filtering these signals have a bandwidth that is half the bandwidth of the original input signal. Therefore both are subsampled by a factor of 2, resulting in the signals $x_0(m)$ and $x_1(m)$ with half the number of samples

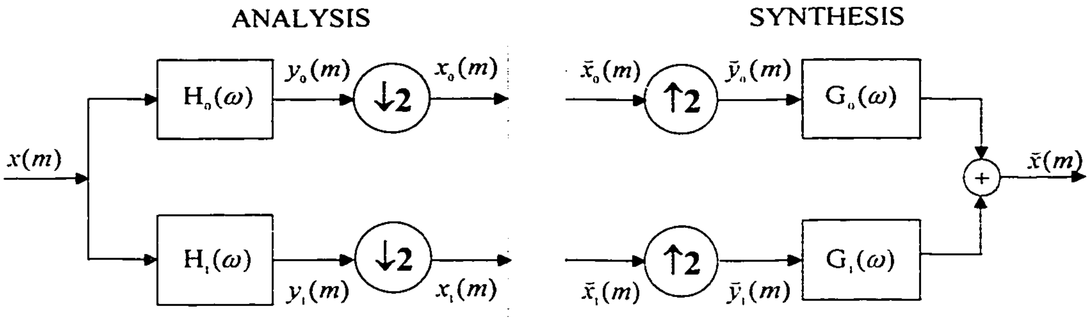


Figure 4-3: Generic two-channel filter bank.

of the input signal $x(m)$. After the processing of the coefficients, the input signal is reconstructed from the two channel signals $\check{x}_0(m)$ and $\check{x}_1(m)$ by first upsampling them to full bandwidth signals $\check{y}_0(m)$ and $\check{y}_1(m)$, then filtering them respectively by the synthesis lowpass and highpass filters $G_0(\omega)$ and $G_1(\omega)$. Finally the filter outputs are added to form the reconstructed signal $\check{x}(m)$.

4.2.2 Perfect reconstruction

Input-output relation

Due to the presence of the *downsampling* and *upsampling* operators the sampling rates within the entire process of analysis and synthesis are half the sampling rate of the input signal. The two-channel filter bank is often referred to as *critically sampled* multirate filter bank. The downsampling or decimation operator is linear but *not* shift invariant. It consists of taking every second sample from the input sequence $x(m)$ as

$$y(m) = x(2m) \tag{4.20}$$

In the frequency domain[2], equation (4.20) translates to

$$Y(\omega) = \frac{1}{2}\left[X\left(\frac{\omega}{2}\right) + X\left(\frac{\omega}{2} + \pi\right)\right] \tag{4.21}$$

where the second term is an indicator of the presence of an *aliasing error*. Note that in the very particular case where the input signal has a bandwidth of only $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, no aliasing will occur.

Upsampling or interpolation of a signal is done by inserting a zero between every sample of the subband signal as

$$\begin{aligned} x(2m) &= y(m) \\ x(2m+1) &= 0 \end{aligned} \tag{4.22}$$

where the frequency domain equivalent is given by

$$X(\omega) = X(2\omega) \tag{4.23}$$

Since $X(\omega)$ has replicas or images of the original spectrum, the upsampler causes an imaging effect that is usually removed by a subsequent interpolating filter.

---

[2] The notation $X(\omega)$ for the *Fourier* transform is used.

Using the above relationships the input-output relationship of the two-channel filter bank can be derived as follows: first the output of the analysis filters is expressed as,

$$\begin{aligned} X_0(\omega) &= \tfrac{1}{2}\left[H_0\left(\tfrac{\omega}{2}\right)X\left(\tfrac{\omega}{2}\right) + H_0\left(\tfrac{\omega}{2}+\pi\right)X\left(\tfrac{\omega}{2}+\pi\right)\right] \\ X_1(\omega) &= \tfrac{1}{2}\left[H_1\left(\tfrac{\omega}{2}\right)X\left(\tfrac{\omega}{2}\right) + H_1\left(\tfrac{\omega}{2}+\pi\right)X\left(\tfrac{\omega}{2}+\pi\right)\right] \end{aligned} \tag{4.24}$$

then these *subband* signals are interpolated as

$$\begin{aligned} \check{Y}_0(\omega) &= \check{X}_0(2\omega) \\ \check{Y}_1(\omega) &= \check{X}_1(2\omega) \end{aligned} \tag{4.25}$$

In order to bring back each subband signal to its original spectral position, the synthesis filters are applied and then the results are added to obtain the reconstructed signal

$$\check{X}(\omega) = G_0(\omega)\check{X}_0(2\omega) + G_1(\omega)\check{X}_1(2\omega) \tag{4.26}$$

Substituting for $\check{X}_0$ and $\check{X}_1$, the following input-output description is obtained

$$\begin{aligned} \check{X}(\omega) \quad = \quad & \frac{1}{2}G_0(\omega)\left[H_0(\omega)X(\omega) + H_0(\omega+\pi)X(\omega+\pi)\right] \\ & + \frac{1}{2}G_1(\omega)\left[H_1(\omega)X(\omega) + H_1(\omega+\pi)X(\omega+\pi)\right] \end{aligned} \tag{4.27}$$

which can be rearranged by grouping the terms related to the input spectrum and those related to the shifted versions as

$$\begin{aligned}\check{X}(\omega) &= \frac{1}{2}\left[G_0(\omega)H_0(\omega) + G_1(\omega)H_1(\omega)\right]X(\omega) \\ &+\frac{1}{2}\left[G_0(\omega)H_0(\omega+\pi) + G_1(\omega)H_1(\omega+\pi)\right]X(\omega+\pi) \end{aligned} \tag{4.28}$$

The reconstructed signal is then a function of a distortion term and an alias term, represented respectively by the first and second lines of equation (4.28). The perfect reconstruction occurs when

$$\begin{aligned} G_0(\omega)H_0(\omega) + G_1(\omega)H_1(\omega) &= 2 \quad \text{distortion factor} \\ G_0(\omega)H_0(\omega+\pi) + G_1(\omega)H_1(\omega+\pi) &= 0 \quad \text{alias gain} \end{aligned} \tag{4.29}$$

Alias cancellation

In order to enforce the perfect reconstruction property for the filter bank one should first get rid of the *alias gain* by specifying the following set of filters

$$\begin{aligned} G_0(\omega) &= 2H_1(\omega+\pi) \\ G_1(\omega) &= -2H_0(\omega+\pi) \end{aligned} \tag{4.30}$$

or equivalently in the z-domain

$$\begin{aligned} G_0(z) &= 2H_1(-z) \\ G_1(z) &= -2H_0(-z) \end{aligned} \tag{4.31}$$

where the factor 2 has been deliberately introduced to remove the $\frac{1}{2}$ in equation (4.28). Finally, the two-channel filter bank becomes a linear and shift invariant system described by the *transfer function*,

$$T(\omega) = \frac{\check{X}(\omega)}{X(\omega)} = H_0(\omega)\, H_1(\omega + \pi) - H_1(\omega)\, H_0(\omega + \pi) \tag{4.32}$$

Distortion cancellation

In the ideal case, $T(\omega)$ is a pure delay, i.e. $T(\omega) = e^{-i\omega\tau}$. This means that the reconstructed signal is a delayed version of the input signal, or $\check{x}(m) = x(m + \tau)$. However, in the general case, $T(\omega)$ represents a distortion and is therefore called the *distortion transfer function* with an amplitude distortion factor $|T(\omega)|$ and a phase distortion factor $\arg\{T(\omega)\}$. It will be shown how the choice of the highpass filter $H_1(\omega)$ can lead to the suppression of phase errors with a small amplitude distortion using the *quadrature mirror filters* (QMF) [39] or to the suppression of both using *conjugate quadrature filters* (CQF) [13]. Finally, the more general choice of *linear-phase biorthogonal perfect reconstruction* filter bank will be discussed.

Quadrature mirror filters (QMF) - The QMF approach consists of choosing the analysis filters according to

$$\begin{array}{lclcl} H_0(\omega) = H(\omega) & \multirow{2}{*}{$\Longleftrightarrow$} & H_0(z) = H(z) & \multirow{2}{*}{$\Longleftrightarrow$} & h_0(l) = h(l) \\ H_1(\omega) = H(\omega + \pi) & & H_1(z) = H(-z) & & h_1(l) = (-1)^l h(l) \end{array} \tag{4.33}$$

where $H(.)$ is a lowpass filter to be determined.

The appellation quadrature mirror filters stems from the fact that the response of $H_1(\omega)$ is the mirror image of the response of $H_0(\omega)$, with respect to the frequency $\frac{\pi}{2}$. The transfer function (4.32) becomes then

$$T(\omega) = H(\omega)^2 - H(\omega + \pi)^2 \tag{4.34}$$

One limitation of the QMF is that either the phase distortion $\arg\{T(\omega)\}$ or the amplitude distortion $|T(\omega|$ can be eliminated, but not both. In the case a small phase distortion is tolerated, it is possible to realize $|T(\omega| = 1$ by using *infinite impulse response* (IIR) filters.

To remove the phase distortion, we use in general QMF that are symmetrical, i.e. linear phase FIR filters of order $L - 1$, having L filters taps, and which are given by

$$H(\omega) = |H(\omega)| \, e^{-i\omega\frac{L-1}{2}} \Longleftrightarrow h(l) = h(L-1-l), \qquad l = 0, \ldots, L-1 \tag{4.35}$$

This choice yields

$$T(\omega) = \left[|H(\omega)|^2 - (-1)^{L-1} |H(\omega + \pi)|^2 \right] e^{i\omega(L-1)} \tag{4.36}$$

Note that choosing L to be *odd* yields $T(\omega) = 0$ for $\omega = \frac{\pi}{2}$, which is a serious amplitude distortion. Therefore L must be *even* leading thus to the following transfer function for the FIR QMF filter bank

$$T(\omega) = \left[|H(\omega)|^2 + |H(\omega + \pi)|^2 \right] e^{i\omega(L-1)} \tag{4.37}$$

which reduces to a pure delay when

$$|H(\omega)|^2 + |H(\omega+\pi)|^2 = 1 \Longleftrightarrow H(z)^2 + H(-z)^2 = 1 \tag{4.38}$$

Conjugate quadrature filters (CQF) - The conjugate quadrature filters[3][13] solution consists in the following choice of analysis filters,

$$\begin{aligned} H_0(\omega) &= H(\omega) & & h_0(l) = h(l) \\ & & \Longleftrightarrow & \\ H_1(\omega) &= -H(-\omega+\pi)e^{-i\omega(L-1)} & & h_1(\omega) = (-1)^l h(L-1-l) \end{aligned} \tag{4.39}$$

or equivalently in the z-domain

$$\begin{aligned} H_0(z) &= H(z) \\ H_1(z) &= -H(-z^{-1})z^{-(L-1)} \end{aligned} \tag{4.40}$$

This choice leads to an exact perfect reconstruction without the use of IIR filters. Moreover using the anti-aliasing choice (4.31), the synthesis filters can be expressed as

$$\begin{aligned} G_0(z) &= 2H_1(-z) = 2H_0(z^{-1})z^{-(L-1)} \\ G_1(z) &= -2H_0(-z) = 2H_1(z^{-1})z^{-(L-1)} \end{aligned} \tag{4.41}$$

where L is even. This results in a practical relationship between the analysis and synthesis filters illustrated by figure 4-4 for the orthogonal *Daubechies* filters with two vanishing moments, i.e. $L = 4$.

---

[3] also called *Smith-Barnwell* filters.

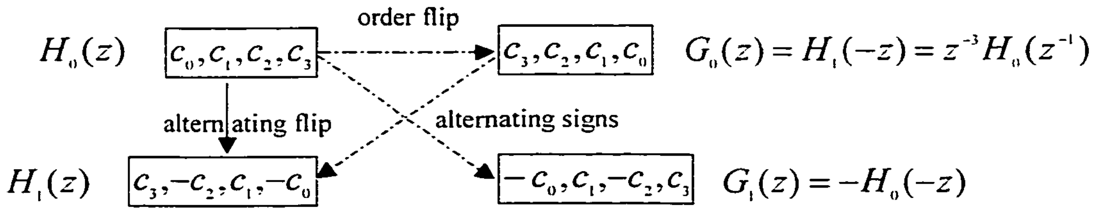


Figure 4-4: Relationships between the orthogonal FIR CQF filter bank illustraded with the Daubechies wavelet system with two vanishing moments.

Remark 5 *The alternating flip automatically gives double-shift orthogonality between the highpass and the lowpass filters. Consequently, when the design of H_0 leads to a perfect reconstruction in the alternating flip filter bank, it also leads to orthogonality. As shown in figure 4-4. the orthogonal Daubechies filters derived in the previous chapter fit this pattern.*

Linear-phase biorthogonal filter bank - The biorthogonal perfect reconstruction condition requires that the product $G_0(z)H_0(z)$ is a *halfband* filter with an odd number of coefficients. Moreover, the center coefficient must be 1. Biorthogonality is a consequence of the cancellation of the aliasing by the dual relations of G_0 to H_1 and G_1 to H_0. Unlike the orthogonal filter bank, the filters lengths can be either *all even* or *all odd.* Thus the analysis filters can be both symmetric, i.e. of odd length, or one symmetric and the other anti-symmetric, i.e. of even length. To cancel aliasing, there is a sign alternation in $G_0(z) = H_1(-z)$ and in $G_1(z) = -H_0(-z)$, where the minus sign does not change the symmetry type. The other possibilities are excluded by the perfect reconstruction condition.

Remark 6 *Note that equation (4.38) is reminiscent of the quadratic condition (3.10).*

However, the wavelet filters design is more flexible than the design of filter bank QMF. Elsewhere, the filter bank filters doe not satisfy any regularity condition, which turns out to be of the outmost importance in the case of iterated filter banks *discussed next.*

4.2.3 Iterated filter banks

From the above we can think of the analysis filters H_0 and H_1 to be the scaling and the wavelet filters respectively. Indeed in the discrete case, i.e. when the input signal is discretized, the MRA structurally defines an *iterated non-uniform*[4] *filter bank.* However, the filters designed for the DWT are regular and lead to a continuous-time scaling and wavelet functions. On the contrary, the filters designed for filter banks, e.g. [51], do not.

We have previously defined the MRA in terms of filtering, where the projections of a signal $f(t)$ onto the scale spaces $P_{V_j}f$ represent the lowpass components of the signal, whereas the projections on the wavelet spaces $P_{W_j}f$ represent the bandpass components. Intuitively, the lowpass components give a multiresolution representation of the signal $f(t)$ at different scales 2^j or resolutions 2^{-j}. The bandpass components as their part, give the details lost when going from a given scale to a coarser one. Otherwise stated, if we rename the projection operators as,

$$H_0^j(f) \equiv f^j = P_{V_j}f \quad \text{and} \quad H_1^j(f) \equiv g^j = P_{W_j}f \tag{4.42}$$

---

[4] The wavelet packet transform is implemented using a uniform filter bank.

then the multiresolution decomposition (2.47) reads,

$$f = H_0^{J_0}(f) + H_1^{J_0}(f) + H_1^{J_0-1}(f) + ... + H_1^{J_1+2}(f) + H_1^{J_1+1}(f) \tag{4.43}$$

which means that the signal $f(t)$ can be recovered starting from the lowest resolution component (i.e. in a scale that does not contain any relevant details) and by successively adding the higher resolution components up to the finest scale.

Elsewhere, from equations (2.29) and (2.38), we know that the projection operators $H_0^j(f)$ and $H_1^j(f)$ give the approximation and detail wavelet coefficients a_k^j and d_k^j, initially defined by the inner products (2.30) and (2.39) but calculated recursively by convolutions (4.13) and (4.15) respectively. Elsewhere, the latter can be expressed as

$$a_k^j = (\downarrow 2)[a_n^{j-1} * c_{-n}^*] \tag{4.44}$$

and

$$d_k^j = (\downarrow 2)[a_n^{j-1} * b_{-n}^*]. \tag{4.45}$$

which is precisely what a filter bank performs, i.e. a discrete convolution followed by downsampling by the factor 2. Consequently, the decomposition *Mallat* algorithm of figure 4-1 is in commonly implemented using the *iterated non-uniform filter bank* of figure 4-5.

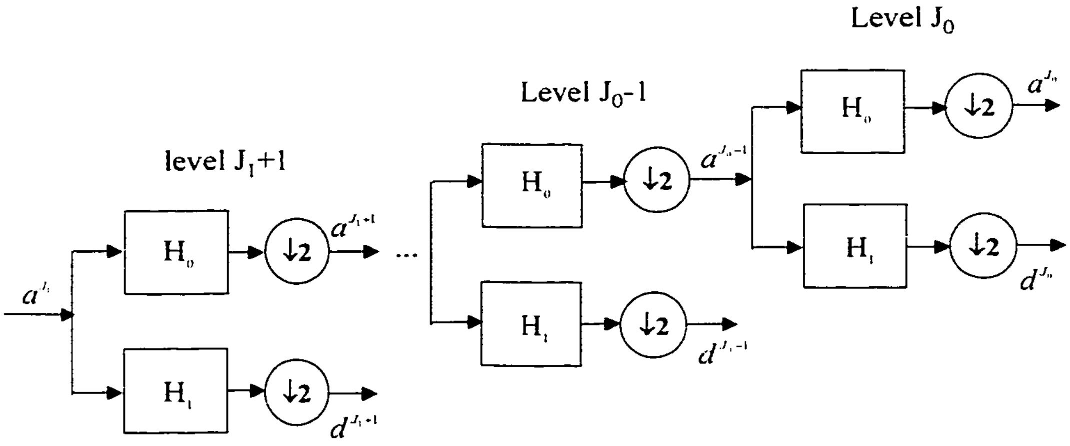


Figure 4-5: Iterated non-uniform filter bank.

4.2.4 Matrix representation

The *Mallat* algorithm is a fast way to compute the DWT using iterated non-uniform filter banks. However, it may be also useful to mathematically represent the DWT or the filter bank as a linear transformation, i.e. a matrix multiplication of the input signal f with a wavelet matrix W.

As an example, we shall compute the matrix W for the *Daubechies* wavelet with two vanishing moments assuming an input signal a^{J_1} consisting of l samples. The convolution of the input sequence with the scaling and wavelet filters is respectively given by the two following circulant matrices,

$$H_0 = \begin{bmatrix} c_0 & c_1 & c_2 & c_3 & & & & & & & \\ & c_0 & c_1 & c_2 & c_3 & & & & & & \\ & & c_0 & c_1 & c_2 & c_3 & & & & & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ & & & & & & & c_0 & c_1 & c_2 & c_3 \\ c_3 & & & & & & & & c_0 & c_1 & c_2 \\ c_2 & c_3 & & & & & & & & c_0 & c_1 \\ c_1 & c_2 & c_3 & & & & & & & & c_0 \end{bmatrix} \in \mathbb{R}^{lxl} \tag{4.46}$$

$$H_1 = \begin{bmatrix} c_3 & -c_2 & c_1 & -c_0 & & & & & & & \\ & c_3 & -c_2 & c_1 & -c_0 & & & & & & \\ & & c_3 & -c_2 & c_1 & -c_0 & & & & & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ & & & & & & & c_3 & -c_2 & c_1 & -c_0 \\ -c_0 & & & & & & & & c_3 & -c_2 & c_1 \\ c_1 & -c_0 & & & & & & & & c_3 & -c_2 \\ -c_2 & c_1 & -c_0 & & & & & & & & c_3 \end{bmatrix} \in \mathbb{R}^{lxl} \tag{4.47}$$

where the blank entries signify zeros.

The downsampling operation is represented by the matrix

$$D = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \in \mathbb{R}^{\frac{l}{2}xl} \tag{4.48}$$

where it has tacitly been assumed that l is even. The combined action of lowpass filtering and downsampling is represented then by the matrix

$$C = DH_0 = \begin{bmatrix} c_0 & c_1 & c_2 & c_3 & & & & & & & \\ & & c_0 & c_1 & c_2 & c_3 & & & & & \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ & & & & & & & c_0 & c_1 & c_2 & c_3 \\ c_2 & c_3 & & & & & & & & c_0 & c_1 \end{bmatrix} \in \mathbb{R}^{\frac{l}{2}xl} \tag{4.49}$$

Similarly, the combined action of highpass filtering and downsampling is given by

$$B = DH_1 \begin{bmatrix} c_3 & -c_2 & c_1 & -c_0 & & & & & & & \\ & & c_3 & -c_2 & c_1 & -c_0 & & & & & \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ & & & & & & & c_3 & -c_2 & c_1 & -c_0 \\ c_1 & -c_0 & & & & & & & & c_3 & -c_2 \end{bmatrix} \in \mathbb{R}^{\frac{l}{2}xl} \tag{4.50}$$

Finally, the wavelet matrix is obtained by combining the two channels of the filter bank as

$$W = \left[\begin{array}{ccccccccccc}
c_0 & c_1 & c_2 & c_3 & & & & & & & \\
c_3 & -c_2 & c_1 & -c_0 & & & & & & & \\
 & & c_0 & c_1 & c_2 & c_3 & & & & & \\
 & & c_3 & -c_2 & c_1 & -c_0 & & & & & \\
 & & & & & & \cdots & & & & \\
 & & & & & & & c_0 & c_1 & c_2 & c_3 \\
 & & & & & & & c_3 & -c_2 & c_1 & -c_0 \\
c_2 & c_3 & & & & & & & & c_0 & c_1 \\
c_1 & -c_0 & & & & & & & & c_3 & -c_2
\end{array}\right] \in \mathbb{R}^{lxl} \qquad (4.51)$$

Remark 7 *A word about the structure of this matrix. The first row generates one component of the data convolved with the scaling filter components* $c_0 \ldots, c_3$. *Likewise the third, the fifth and the other odd rows. If the even rows followed this pattern offset by one, then the matrix would be a* circulant, *that is an ordinary convolution that could be done using FFT methods. However, the even rows perform a different convolution with wavelet filters* $c_3, -c_2, c_1, -c_0$ *obtained by the alternating flip as exemplified by figure 4-4. Thus the overall action of the matrix* W *is to perform two related convolutions, then to decimate each of them by* 2 *and interleave the remaining halves.*

The DWT consists then in applying the wavelet matrix W (4.51) hierarchically, first to the full data vector of length l, then after a permutation, to the *smooth* (lowpass filtered)

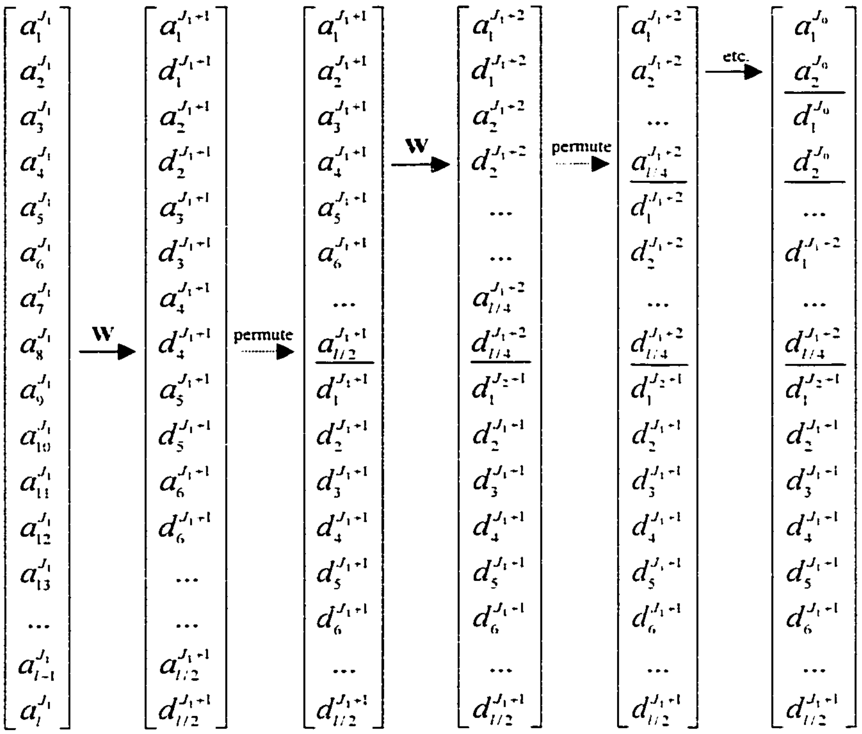


Figure 4-6: Schematic representation of the DWT. At each stage of the transform, the input vector is multiplied by the wavelet matrix W then the approximation and detail coefficients are rearranged by permutation such that the second pass of the DWT is performed on the approximation coefficients only.

vector of length $l/2$, then to the *smooth-smooth* vector of length $l/4$, and so on until a trivial number of *smooth-...-smooth* components, usually 2, remain. The DWT consists of these remaining components and all the details coefficients that were accumulated along the way. This is illustrated by the diagram of figure 4-6.

Using the matrix notation, the input data can be reconstructed by applying the inverse wavelet matrix W^{-1} to the wavelet coefficients vectors. It is clear that the perfect reconstruction property of the two-channel filter bank $W^{-1} \circ W$ is related to the invertibilty of the matrix W. In the case of conjugate quadrature filters (CQF) the matrix W

is orthogonal and $W^{-1} = W^t$ which shows the importance of orthogonal wavelets in the simplification of the computation load. Finally, notice from the matrix (4.51) how the last two rows warp around like convolutions with periodic boundary conditions. The point is elaborated in the sequel.

4.3 Boundaries Treatments

Seismic data do not extend indefinitely in time or space, but is limited to a finite interval fixed by the number of CMP and the TWT in the x and y directions respectively. However, the theoretical DWT operator $W_{l=\infty}$ (4.51) maps an infinite-dimension signal $f(t)$ to an infinite set of wavelet coefficients. To apply W to finite length signals like seismic data, we need either to truncate W to a finite invertible operator or to extend the signal to an infinite one. The resulting transform differs at the boundaries, while remaining the same for data far enough from the boundaries. Truncating the wavelet operator means designing the so-called *wavelets on the interval* [18], which have proven to be effective in theory but are not satisfactory from a practical point of view [68]. It is often preferable to use simple schemes based on signal extension at the boundaries. These methods include: *zero-padding*, *periodic* and *symmetric extensions.*

4.3.1 Zero padding

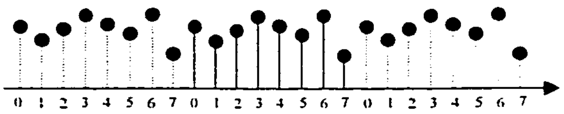
This technique assumes that the signal is zero outside the original finite support, i.e. padding the signal with zeros. This approach is not acceptable in the context of data compression because it yields much more wavelet coefficients than the original data. If one truncates the wavelet transform coefficients to get the same number as data samples, this will distort the signal and jeopardize the perfect reconstruction of the data.

4.3.2 Periodic extension

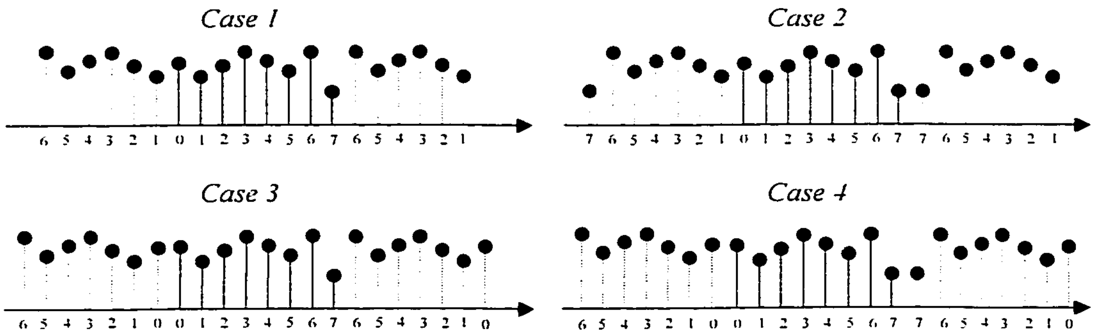
This technique consists of extending the finite signal by periodic replication as shown in figure 4-7-(a). After a DWT the coefficients that lie outside the interval in which the original signal is defined can be simply discarded. Since these coefficients are the same as those retained, the original signal can be recovered without any distortion problems. However, unless the first and the last samples have the same value, unwanted discontinuity are introduced at the boundaries. The consequent detail coefficients will have large value and may deteriorate the performance of the compression.

4.3.3 Symmetric extension

A better solution is the symmetric extension of the original signal. This is done by mirroring the signal around its endpoints, which makes the discontinuities disappear but not their higher order derivatives. Symmetric extension can be thought of as a form of periodic extension applied to a concatenation of the original signal and a mirrored copy of it. After filtering, one can retain twice as many coefficients. Fortunately one can discard half of them if the filters are symmetric, yielding the same number of coefficients as the original data. This emphasizes the importance of using the symmetric biorthogonal wavelets. Moreover there are four possible ways to symmetrically extend a discrete signal as depicted by figure 4-7-(b).



(a) Periodic extension



(b) Symmetric extension

Figure 4-7: Boundaries treatments.

4.4 Extension to 2-D Signals

For 2-D data, it is common to use separable implementations where filter banks are applied to each dimension individually. There are two common ways to generalize the 1-D DWT. The first transform leads to the so-called *standard decomposition.* As illustrated in figure 4-8 for a $4x4$ input matrix data and using the *Haar* transform, it consists of applying first the 1-D DWT to each row of the input data, e.g. pixels in the case of digital images. This operation gives an average value along with detail coefficients for each row. Next, we treat these transformed rows as if they were themselves an image and apply the 1-D to each column. The resulting values are all details coefficients except for a single overall average coefficient, i.e. $a^{2c}_{a^{2r}}$.

The second type of separable 2-D DWT is called the *non-standard decomposition.*

For the sake of comparison, it is also shown in figure 4-9 for the same input data and transform. Note how it alternates between operations on rows and columns, by first performing one step of pairwise averaging and differencing on the pixel values in each row of the image matrix. Next, a similar pairwise vertical averaging and differencing is applied to each column of the result. To complete the transformation, the same process is repeated recursively only on the quadrant containing averages in both directions, i.e. the submatrix,

$$\begin{bmatrix} a^{1c} & a^{1c} \\ a^{1c} & a^{1c} \end{bmatrix} \tag{4.52}$$

In the subband parlance, this matrix is called the *low-low (or residual) subband* as it results from the lowpass filtering in both directions of the 2-D data. Indeed, at each stage of the decomposition, three more subbands are generated: The *high-low (or horizontal) subband* consists of highpass filtering in the rows direction followed by a lowpass filtering in the columns direction. This subband contains the wavelet coefficients emphasizing the horizontal details of the data. The *low-high (or vertical) subband* is the dual of the high-low subband and contains the vertical detail wavelet coefficients. Finally the *high-high (or diagonal) subband* consists of highpass filtering in both the directions and gives the diagonal detail wavelet coefficients. In the course of the present dissertation, we use the non-standard decomposition, whose 2-D bases are given by equations (3.61), (3.63), (3.64), and (3.65). Indeed the nonstandard decomposition is more efficient as the DWT is iterated on subimages rather than on the full dimension image. In figure 4-10-(a) we apply a one-

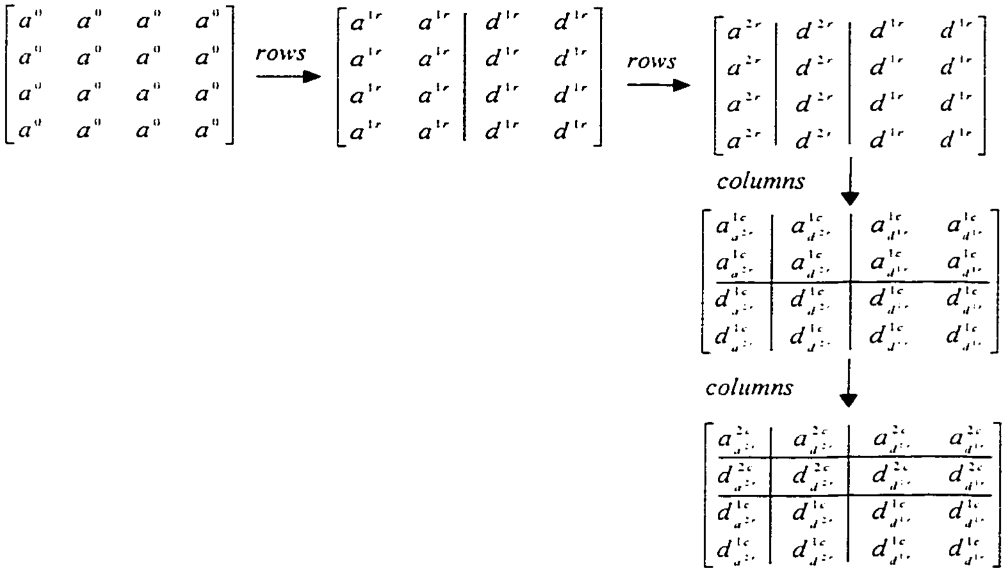


Figure 4-8: Schematic representation of the standard decomposition of 2-D data matrix using the *Haar* DWT.

stage non-standard 2-D DWT using the two-channel filter bank defined earlier to a seismic image using the *Daubechies* wavelet system with two vanishing moments. The division of the frequency spectrum is depicted in figure 4-10-(b), but the wavelet coefficients are in general, arranged in the *Mallat format* as shown in 4-11. This format is more suitable for the processing of the wavelet coefficients.

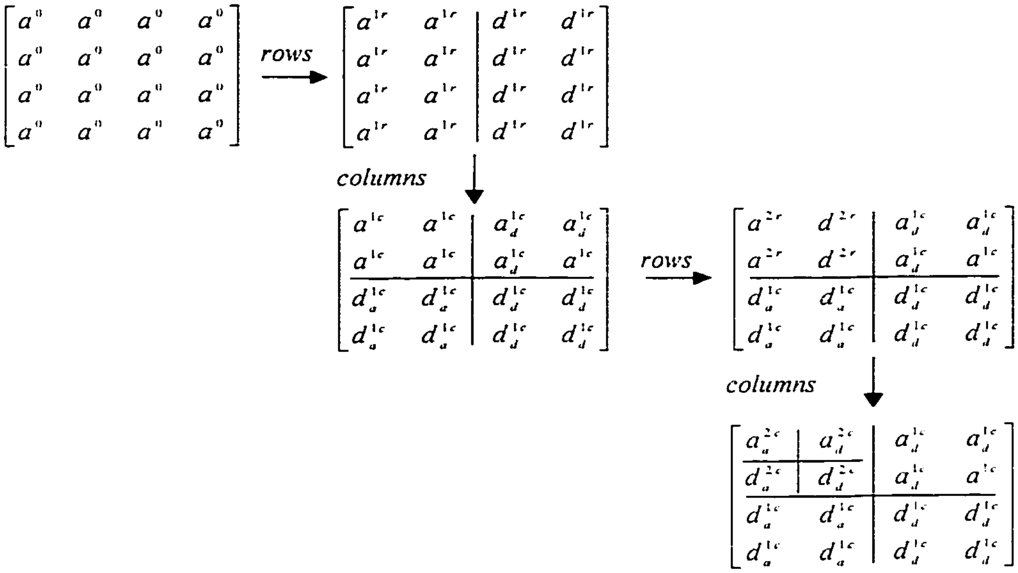


Figure 4-9: Schematic representation of the non-standard decomposition of 2-D data matrix using the *Haar* DWT.

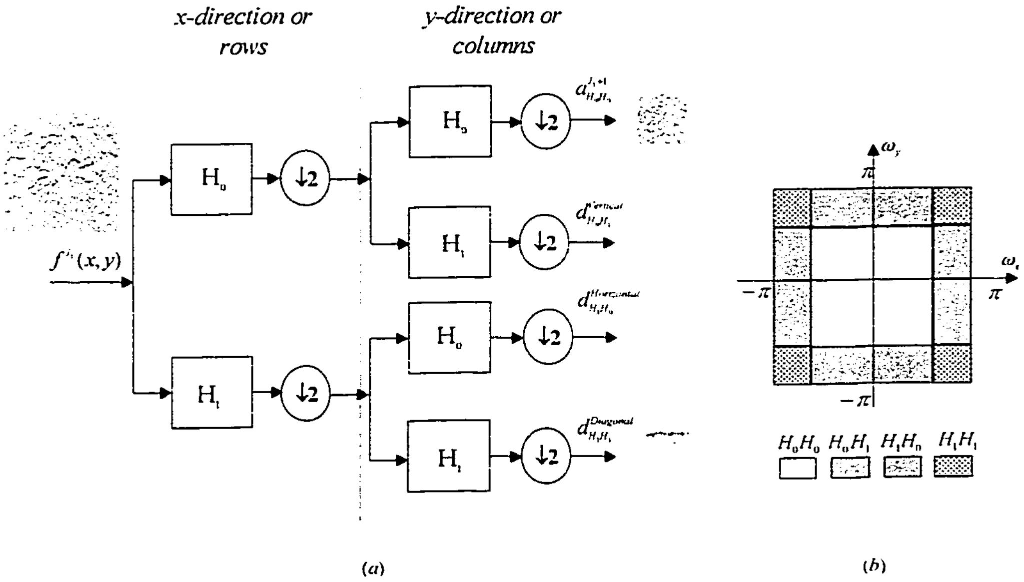


Figure 4-10: (a) One-stage non-standard 2-D DWT implemented using a two-channel filter bank. (b) Frequency spectrum division.

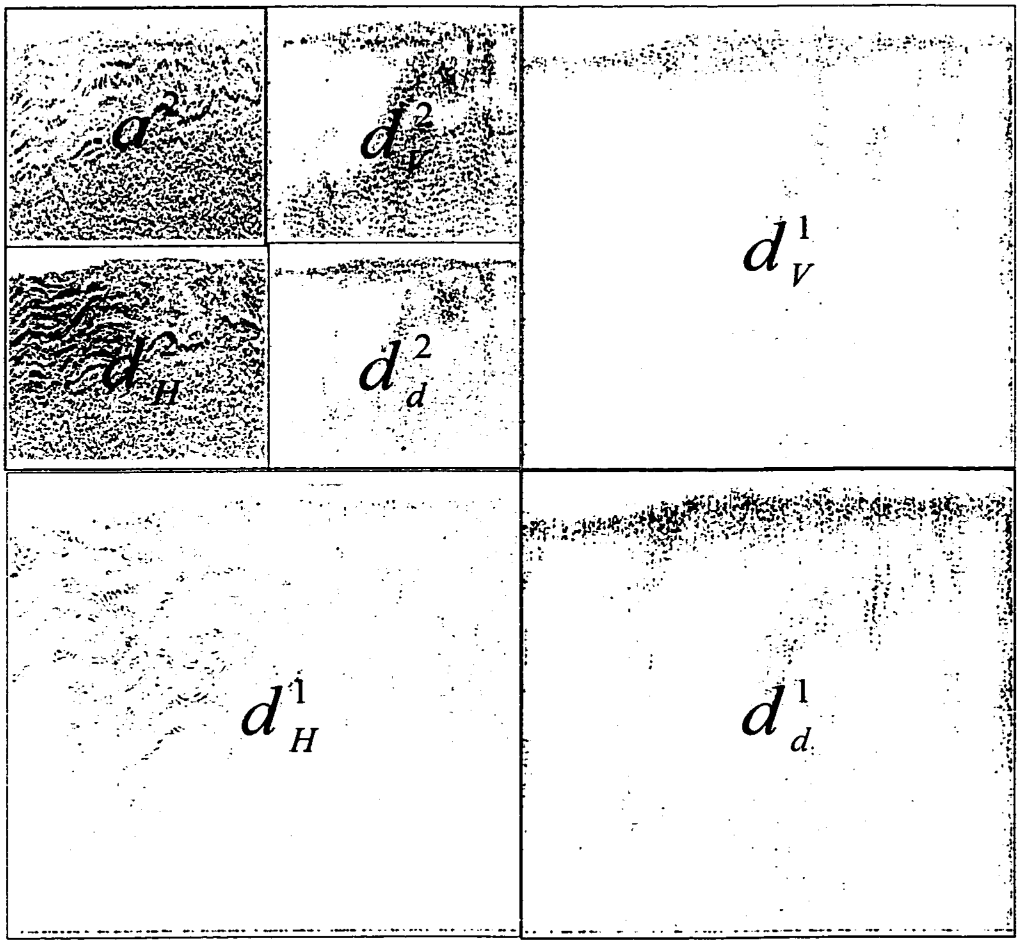


Figure 4-11: A two-level nonstandard wavelet decomposition of a seismic image.

Chapter 5

LIFTED WAVELET TRANSFORM

The lifting scheme is a flexible and powerful algorithm for constructing wavelets entirely in the spatial domain. Therefore, it is ideally suited to build *second generation wavelets*, i.e. wavelets adapted to situations that do not allow translations and dilations. These could be wavelets on bounded domains, wavelets on curves and surfaces, weighted wavelets, and wavelets of irregular sampled data [96]. In the case of the *first generation wavelets*, i.e. the translation and dilation invariant case, the lifting scheme leads to an efficient factorization of the DWT filter bank into cascaded cells known as *ladder* structures [54]. Furthermore, the lifting scheme yields the fastest implementation of the DWT, a fully in-place calculation and a straightforward inverse transform.

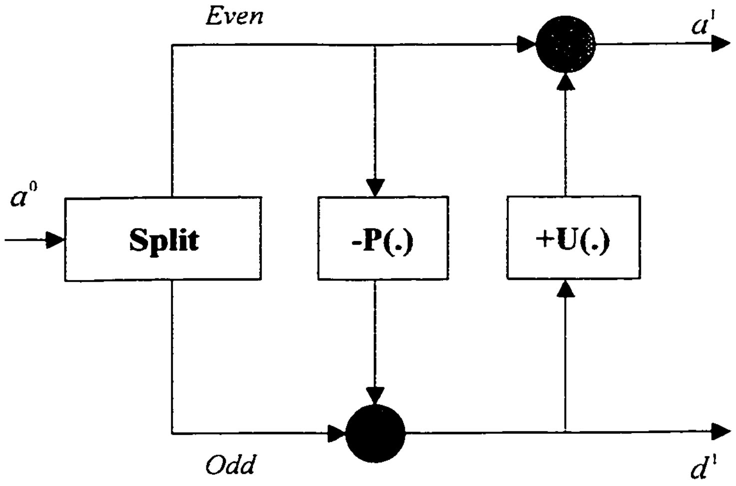


Figure 5-1: One-stage forward lifting scheme.

5.1 Lifting Scheme

5.1.1 Lifting scheme machinery

The lifting scheme consists of three consecutive operators arranged as in figure 5-1. These are:

- **Split** - Decorrelates the original data set by partitioning it into two disjoint subsets. A trivial but very interesting split operator is the *lazy wavelet transform* that divides the input data[1] a_k^0 into the even indexed samples a_{2k}^0 and the odd indexed ones a_{2k+1}^0.
- **Predict** - Generates through *dual lifting*, the detail (wavelet) coefficients d_k^1 as the error in predicting odd subset from the even one, i.e. $d_k^1 = a_{2k+1}^0 - P(a_{2k}^0)$, where $P(.)$ is an arbitrary prediction operator.

---

[1] we keep the same terminlogy and indexing as for the DWT.

- **Update** - Restores, through *primal lifting*, the mean affected by the aliasing introduced by the lazy transform. The even indexed samples that represent a coarse approximation of the data at a higher scale, are *lifted* by an update operator $U(.)$ as $a_k^1 = a_{2k}^0 + U(d_k^1)$, which yields the approximation coefficients a_k^1.

5.1.2 Lifting in the space domain

In the space domain, the lifting scheme can be viewed as a *prediction-error* decomposition. The approximation coefficients a_k^1 are predictors at the next higher scale for the input data a_k^0. The detail coefficients d_k^1 are then the prediction-errors between the approximation coefficients and the data that they attempt to predict. The essence of the lifting scheme, like other transforms, is to decorrelate the data in view of capturing the information it contains with fewer coefficients than samples in the original data. Such a compact representation produces data compression[2]. In order to get a maximal data reduction from prediction, it is required that the subsets generated by the split operator should be maximally correlated. Thus, the lazy wavelet transform is more convenient because it interlaces the two subsets by splitting the data into even and odd indexed partitions. The predict stage is intended for obtaining a more compact representation of the data. It attempts to predict the odd indexed samples based on the high correlation present in the data and in the even indexed partition. The construction of predictors is based on the concept of *interpolating subdivision scheme* [91]. The latter is characterized by its order N, which sets the smoothness of the interpolating polynomial used to find the wavelet

---

[2] To be distinguished from the equally challenging problem of quantization and encoding dealt with in chapter 6.

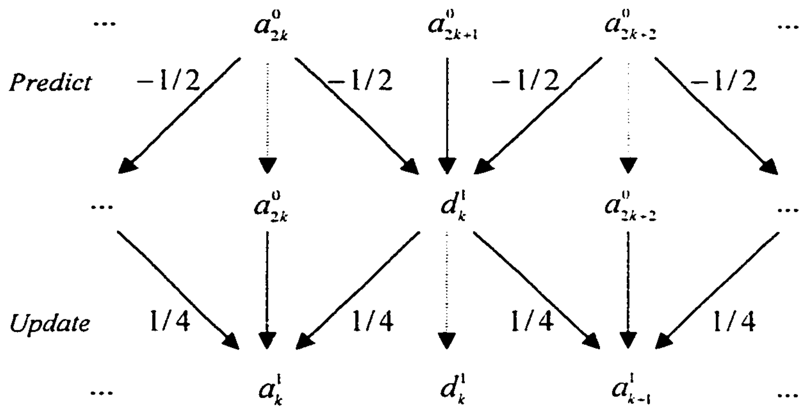


Figure 5-2: Parallel structure of the forward (2,2)-biorthogonal lifted DWT.

coefficients. This polynomial is referred to as the *dual wavelet* and N as the number of *dual vanishing moments.* For instance, in the case of the *lifted (2,2)-biorthogonal wavelet transform* depicted in figure 5-2, an odd sample is predicted as the average of its two even neighbors, i.e.

$$d^1_k = a^0_{2k+1} - \frac{1}{2}(a^0_{2k} + a^0_{2k+2}) \tag{5.1}$$

The model of the predictor $P(.)$ is thus a piecewise linear function over intervals of length two and the detail coefficients d^1_k measure to which extent the original data fails to be linear. Although the even indexed samples represent a low-resolution smooth version of the original data, they suffer from the loss of some statistical properties, e.g. the *mean.* The update stage consists then in finding a scaling function[3] using the previously calculated wavelet coefficients in order to preserve the mean or to cancel the aliasing effect

---

[3] In anticipation, we can say that the update operator mimics the scaling function, whereas the predict operator mimics the wavelet.

of the subsampling, among all the approximation coefficients a^j throughout all levels. This function will have an order depending of some even value $\tilde{N}$ referred to as the number of *real vanishing moments.* The rationale of the update stage is that each wavelet coefficient d_k^1 will give back to the approximation coefficient a_k^0 used to predict it as much as it received, and this amount is given by the lifting coefficients. Again, in the case of the *(2,2)-biorthogonal wavelet transform*, the lifting coefficients are $\left\{\frac{1}{4}, \frac{1}{4}\right\}$ and the update stage can be expressed as [22],

$$a_k^1 = a_{2k}^0 + \frac{1}{4}(d_k^1 + d_{k-1}^1) \tag{5.2}$$

These three steps form a one-stage lifted wavelet transform and the full multiresolution decomposition is then obtained, as in the case of the DWT iterated filter bank, by iterating the primal and dual lifting steps on the scaling coefficients at the next coarser scale.

5.2 Polyphase Filter Banks

In order to factorize the DWT filter bank, it is convenient to express the latter using *modulation matrices*, whose special structure is revealed using the *polyphase form.* This will automatically lead to the *lattice* factorization of *unitary* filter banks. However, in the non-unitary case such a factorization fails [22],[92]. Therefore, the lifting scheme offers a more versatile and broader spectrum factorization into *ladder* structures [54].

5.2.1 Modulation matrices

The perfect reconstruction conditions (4.29) are expressed in the z-domain as

$$\begin{aligned} G_0(z)H_0(z) + G_1(z)H_1(z) &= 2 \\ G_0(z)H_0(-z) + G_1(z)H_1(-z) &= 0 \end{aligned} \tag{5.3}$$

where we assumed without loss of generality that all the filters are FIR and have their coefficients centered around the zero position [92]. Furthermore, equation (5.3) can be factored as,

$$\begin{bmatrix} G_0(z) & G_1(z) \end{bmatrix} \begin{bmatrix} H_0(z) & H_0(-z) \\ H_1(z) & H_1(-z) \end{bmatrix} = \begin{bmatrix} 2 & 0 \end{bmatrix} \tag{5.4}$$

where the $2x2$ matrix $\mathbf{H}_m$ is called the *analysis modulation matrix* and plays an important role in filter bank theory [100]. The *synthesis modulation matrix* is defined as [92],

$$\mathbf{G}_m = \begin{bmatrix} G_0(z) & G_1(z) \\ G_0(-z) & G_1(-z) \end{bmatrix} \tag{5.5}$$

which allows to express the prefect reconstruction conditions in the following matrix form

$$\begin{bmatrix} G_0(z) & G_1(z) \\ G_0(-z) & G_1(-z) \end{bmatrix} \begin{bmatrix} H_0(z) & H_0(-z) \\ H_1(z) & H_1(-z) \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \tag{5.6}$$

5.2.2 Polyphase form

The direct implementation of a filter bank, where the downsampling and the upsampling operations occur respectively after and before the filtering operations, is not efficient. For instance, in the analysis stage, we are performing convolution with all samples and then discarding half of them. Nonetheless, if we do not compute the latter, the system is still working even at a faster rate, but the output is at a half rate because of the downsampling. This is the *polyphase form*, which operates on the different phases separately. Indeed, the input vector is separated into even, x^e, and an odd, x^o, parts such that each phase of the subband filters acts simultaneously on separate phases of the input. For instance, the polyphase representation of the lowpass filter of the filter bank is,

$$\begin{aligned} H_0(z) &= \sum_k c_{2k} z^{-k} + z^{-1} \sum_k c_{2k+1} z^{-k} \\ &= H_0^e(z^2) + z^{-1} H_0^o(z^2) \end{aligned} \tag{5.7}$$

where we factored out z^{-1} to get powers of z^2 for the even part $H_0^e(z^2)$ and the odd part $H_0^o(z^2)$ of the lowpass filter. In order to interchange the filtering and downsampling operations, we use the *Noble identities*, that are outlined in figure 5-3, to redraw the two-channel filter bank as shown in figure 5-4 for the lowpass channel. To extend the polyphase form to the complete two-channel filter bank, it is more convenient to use the polyphase

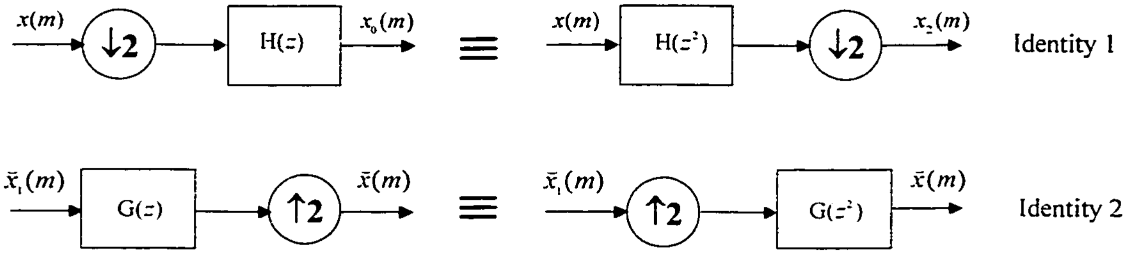


Figure 5-3: The Noble identities.

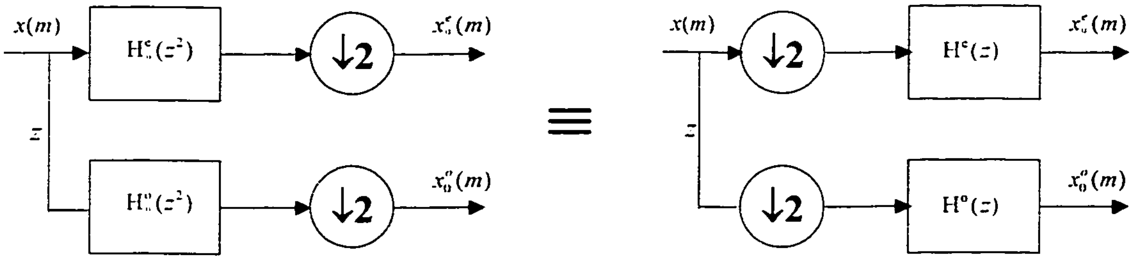


Figure 5-4: Polyphase transformation of the lowpass channel using *Nobel* identity 1.

matrix representation. For, define the *analysis polyphase matrix* as

$$\mathbf{H}_p(z) = \begin{bmatrix} H_0^e(z) & H_0^o(z) \\ H_1^e(z) & H_1^o(z) \end{bmatrix} \tag{5.8}$$

then the subband signals $x_0(m)$ and $x_1(m)$ (we refer to figure 4-3) become,

$$\begin{bmatrix} X_0(z) \\ X_1(z) \end{bmatrix} = \mathbf{H}_p(z) \begin{bmatrix} X^e(z) \\ zX^o(z) \end{bmatrix} \tag{5.9}$$

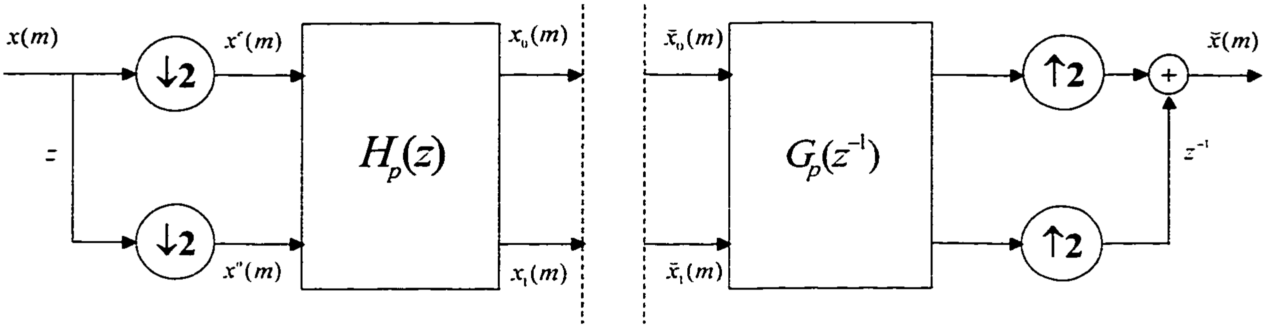


Figure 5-5: Polyphase representation of the two-channel filter bank.

Similarly, we define the *synthesis polyphase matrix* as

$$\mathbf{G}_p(z^{-1}) = \begin{bmatrix} G_0^e(z) & G_1^e(z) \\ G_0^o(z) & G_1^o(z) \end{bmatrix} \tag{5.10}$$

where the time-reversal is required in order to cancel delays caused by the polyphase representation as shown for the one-stage polyphase representation of the fast wavelet transform of figure 5-5.

Remark 8 *The polyphase matrix H_p is connected to the modulation matrix H_m by a 2-point DFT and a diagonal delay matrix as [92]*

$$\begin{bmatrix} H_0^e(z^2) & H_0^o(z^2) \\ H_1^e(z^2) & H_1^o(z^2) \end{bmatrix} = \frac{1}{2} \begin{bmatrix} H_0(z) & H_0(-z) \\ H_1(z) & H_1(-z) \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & z \end{bmatrix} \tag{5.11}$$

Example 4 ***(4-tap Daubechies wavelet system)*** *- Consider the wavelet matrix W (4.51) which, we recall, is the result of interleaving the highpass and the lowpass filter coefficients both downsampled by a factor of* 2.

To avoid the problem of boundaries, we consider the infinite dimension case, i.e.

$$\textit{Block} \quad H_b = \begin{bmatrix} \cdot & \cdot & & & & & & \\ \cdot & \cdot & & & & & & \\ c_0 & c_1 & c_2 & c_3 & & & & \\ c_3 & -c_2 & c_1 & -c_0 & & & & \\ & & c_0 & c_1 & c_2 & c_3 & & \\ & & c_3 & -c_2 & c_1 & -c_0 & & \\ & & & & \cdot & \cdot & \cdot & \cdot \\ & & & & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

where the matrix has been renamed H_b to emphasize its block Toeplitz form. To derive the polyphase form, we separate the even and the odd coefficients as

$$\textit{Polyphase} \quad H_p = \begin{bmatrix} \cdot & & & & \cdot & & & \\ c_0 & c_2 & & & c_1 & c_3 & & \\ & c_0 & c_2 & & & c_3 & c_1 & \\ & & \cdot & \cdot & & & \cdot & \cdot \\ \cdot & & & & & & & \\ c_3 & c_1 & & & -c_2 & -c_0 & & \\ & c_3 & c_1 & & & -c_2 & -c_0 & \\ & & \cdot & \cdot & & & \cdot & \cdot \end{bmatrix} = \begin{bmatrix} H_0^e & H_0^o \\ H_1^e & H_1^o \end{bmatrix}$$

5.2.3 Perfect reconstruction revisited

One reason for introducing the polyphase form is efficiency. The analysis and synthesis banks are faster, when the downsamplers and upsamplers are moved outside. The even and odd subfilters that appear inside are about 50% shorter. So convolutions can be executed more quickly.

The other reason is to simplify the theory of perfect reconstruction filter banks. Indeed, from figure 5-5, we can see that the analysis and the synthesis matrices are side to side. Moreover, those matrices stem from pure time-invariant filters, i.e. the even and odd phases of the scaling and wavelet filters. In between would come compression or denoising as it will be discussed in the next chapter. However, in the absence of any processing, we have

$$\mathbf{G}_p(z^{-1})\mathbf{H}_p(z) = I \tag{5.12}$$

which is the matrix form of the perfect reconstruction conditions (5.3) or (5.6). Elsewhere, if we assume that $\mathbf{H}_p(z)$ is invertible, we find that

$$\mathbf{G}_p(z^{-1}) = \frac{1}{H_0^e(z)H_1^o(z) - H_0^o(z)H_1^e(z)} \begin{bmatrix} H_1^o(z) & -H_0^o(z) \\ -H_1^e(z) & H_0^e(z) \end{bmatrix} \tag{5.13}$$

Moreover, if we impose that the determinant of $\mathbf{H}_p(z)$ is one, then not only will $\mathbf{H}_p(z)$

be invertible, but also

$$
\begin{aligned}
H_0^e(z) &= G_1^o(z^{-1}) \quad ; H_0^o(z) = -G_1^e(z^{-1}) \\
H_1^e(z) &= -G_0^o(z^{-1}) \quad ; H_1^o(z) = G_0^e(z^{-1})
\end{aligned}
\tag{5.14}
$$

i.e. we get a *biorthogonal filter bank*. Note that in the special case $H_0 = G_0$ and $H_1 = G_1$, the polyphase matrix is *unitary*[4] or equivalently the DWT is orthogonal. Elsewhere, the analysis wavelet filters H_0 and H_1 are called *complementary* as well as the synthesis wavelet filters G_0 and G_1. Moreover all the phases pairs are relatively prime. An interesting instance, however is when the polyphase matrix is identity, i.e. $H_0(z) = 1$ and $H_1(z) = z^{-1}$. This is the lazy wavelet transform, which in the filter bank parlance is called the *polyphase transform.*

5.2.4 Lattice factorization

Any unitary DWT two-channel filter bank can be factored into a lattice form, using rotation and delay matrices [92]. The rotations matrices or *Givens* rotations are defined as

$$
R_m = \begin{bmatrix} \cos\theta_m & \sin\theta_m \\ -\sin\theta_m & \cos\theta_m \end{bmatrix} = \cos\theta_m \begin{bmatrix} 1 & k \\ -k & 1 \end{bmatrix}
\tag{5.15}
$$

---

[4] or para-unitary in the case $|z| = 1$.

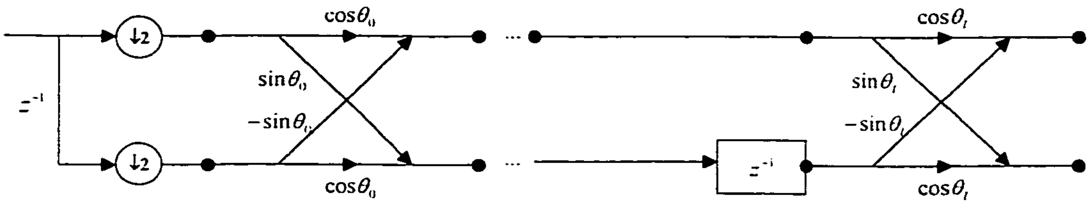


Figure 5-6: Lattice factorization using rotations and delays.

where the design parameter $k = \tan\theta$ is used in the implementation to save multiplications. The delay matrices that are expressed as,

$$\Lambda(z) = \begin{bmatrix} 1 & 1 \\ 0 & z^{-1} \end{bmatrix} \tag{5.16}$$

are used to delay the second channel. It is easy to verify that both the above operators are unitary so that any cascade of the form

$$H_p(z) = R_l(z)\Lambda(z)R_{l-1}(z)\Lambda(z)...R_1(z)\Lambda(z)R_0(z) \tag{5.17}$$

is also unitary and yields the lattice structure of figure 5-6.

Remark 9 *The lattice structure posses a hierarchical property, i.e. if we simply delete one or more cells from the structure, the perfect reconstruction is preserved (as the remainder lattice is still unitary!) This property, however, is not possible if we implement the filters using a direct-form structure [100]. The lattice factorization fails however in the case of non-unitary filters bank, i.e. the biorthogonal DWT.*

5.3 Fast Lifted DWT

The lifting scheme can either be used to gradually build a multiresolution with particular properties or to factorize the FIR filters of an existing DWT into ladder structures. The underlying concepts are common and are summarized by the following two results due to *Sweldens* [94].

5.3.1 Primal and dual lifting

Definition 5 ***(Primal Lifting)*** *- Let* (G_0, G_1) *be complementary. Then any other FIR filter* G_1^{new} *complementary to* G_0 *is of the form*

$$G_1^{new}(z) = G_1(z) + G_0(z)U(z^2) \tag{5.18}$$

where $U(z)$ *is a Laurent polynomial.*

This can be seen readily if we write the new polyphase matrix, after the primal lifting, i.e.

$$\begin{aligned} \begin{bmatrix} G_0^e(z) & G_1^e(z) + G_0^e U(z) \\ G_0^o(z) & G_1^o(z) + G_0^o U(z) \end{bmatrix} &= \begin{bmatrix} G_0^e(z) & G_1^e(z) \\ G_0^o(z) & G_1^o(z) \end{bmatrix} \begin{bmatrix} 1 & U(z) \\ 0 & 1 \end{bmatrix} \\ \mathbf{G}_p^{new}(z^{-1}) &= \mathbf{G}_p(z^{-1}) \begin{bmatrix} 1 & U(z) \\ 0 & 1 \end{bmatrix} \end{aligned} \tag{5.19}$$

where it is straightforward to check that $\left|\mathbf{G}_p^{new}(z^{-1})\right| = 1$. Consequently, (G_0, G_1^{new}) are also complementary. Elsewhere, using the perfect reconstruction property (5.12), we can

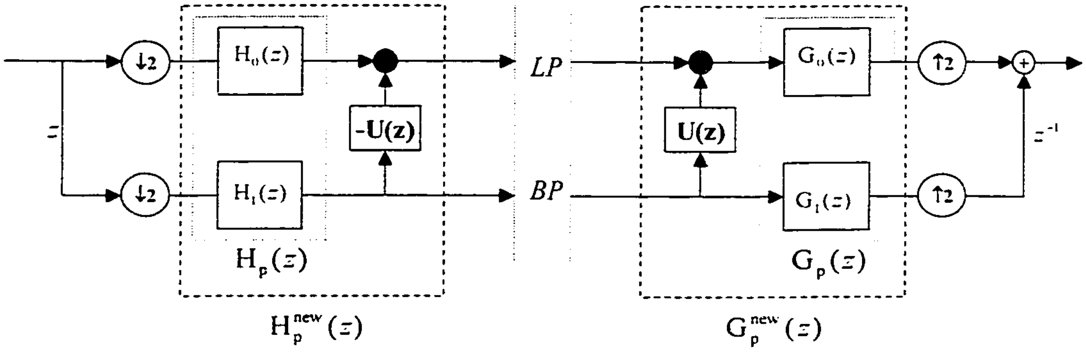


Figure 5-7: The primal lifting: The original filter bank is followed by a lifting cell, which lifts the lowpass subband (LP) with the help of the highpass (or bandpass) subband (BP).

express the new analysis polyphase matrix as

$$\mathbf{H}_p^{new}(z) = \mathbf{H}_p(z) \begin{bmatrix} 1 & 0 \\ -U(z^{-1}) & 1 \end{bmatrix} \tag{5.20}$$

which implies that the primal lifting creates the new lowpass filter

$$H_0^{new}(z) = H_0(z) - H_1(z)U(z^{-2}) \tag{5.21}$$

This expression can be interpreted as lifting the lowpass subband with the help of the highpass subband as illustrated by figure 5-7.

Definition 6 ***(Dual Lifting)*** *- Let (G_0, G_1) be complementary. Then any other FIR filter G_0^{new} complementary to G_1 is of the form*

$$G_0^{new}(z) = G_0(z) + G_1(z)P(z^2) \tag{5.22}$$

where $P(z)$ is a Laurent polynomial.

After dual lifting the new synthesis polyphase matrix is given by

$$\mathbf{G}_p^{new}(z^{-1}) = \begin{bmatrix} G_0^e(z) + G_1^e P(z) & G_1^e(z) \\ G_0^o(z) + G_1^o P(z) & G_1^o(z) \end{bmatrix} = \mathbf{G}_p(z^{-1}) \begin{bmatrix} 1 & 0 \\ P(z) & 1 \end{bmatrix} \tag{5.23}$$

Again, using the perfect reconstruction property, we get

$$\mathbf{H}_p^{new}(z) = \mathbf{H}_p(z) \begin{bmatrix} 1 & -P(z) \\ 0 & 1 \end{bmatrix} \tag{5.24}$$

which implies that the dual lifting creates the new highpass filter

$$H_1^{new}(z) = H_1(z) - H_0 P(z^{-2}) \tag{5.25}$$

Similarly, equation (5.25) can be interpreted as lifting the highpass subband with the help of the lowpass subband as shown in figure 5-8.

The lifting scheme can be used to build wavelet transforms starting from the trivial polyphase transform or the lazy wavelet transform. At each stage, a dual lifting is first applied to realize the space localization of the transform. The frequency localization is achieved, as its part, by the primal lifting. The invertibility of the lifted wavelet transform is preserved at each stage because of the triangular structure of the lifting cells. Moreover, it is important to note that the inverse lifted wavelet transform is simply performed by undoing the lifting steps by applying the same *Laurent* polynomials, $P(z)$ or $U(z)$ as

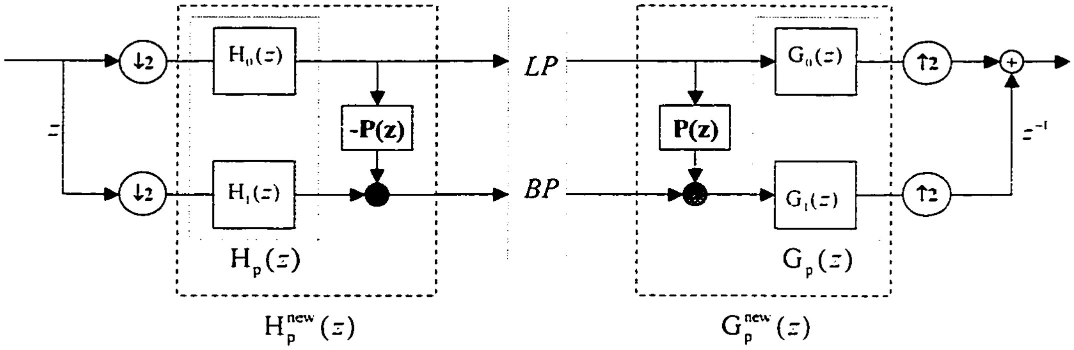


Figure 5-8: The dual lifting: The original filter bank is followed by a lifting cell, which lifts the highpass subband (BP) with the help of the lowpass subband (LP).

illustrated by figures 5-7 and 5-8. The other important feature of the lifted wavelet transform is the in-place calculation of the wavelet coefficients. Indeed, every time we apply a primal or a dual or both lifting steps we replace the input sample stream by the wavelet coefficients without the need for any auxiliary memory.

Finally, to establish the connection between the lifting scheme in the space domain and the polyphase form, note that lifting the highpass subband with the help of the lowpass subband can be seen as predicting the odd samples from the even samples. Thus the dual lifting could be seen as the frequency domain counterpart of the predict stage of the lifting scheme. Similarly, the primal stage is considered as the counterpart of the update stage, because lifting the lowpass subband with the highpass subband is aimed at keeping some statistical properties of the input stream by curing the aliasing introduced by the polyphase representation.

5.3.2 Lifting factorization

In the previous section we showed how to lift a wavelet transform in view of progressively building a more sophisticated multiresolution analysis. We can also use the lifting scheme to factorize any FIR filters of an existing wavelet transform into ladder structures. As a result, the appealing properties of the lifting scheme can be efficiently combined to improve the transform stage of the seismic compression scheme.

Factoring Algorithm

Daubechies and *Sweidens* [22] showed that any synthesis polyphase matrix representing a DWT with FIR filters can be factored into a finite product of unit upper and lower triangular $2x2$ matrices, and a diagonal normalization matrix as

$$\mathbf{G}_p(z^{-1}) = \overbrace{\begin{bmatrix} K & 0 \\ 0 & 1/K \end{bmatrix}}^{\text{normalization}} \prod_{i=1}^{m} \text{predict stage} \underbrace{\overbrace{\begin{bmatrix} 1 & 0 \\ P_i(z) & 0 \end{bmatrix}}^{\text{dual lifting}}} \text{update stage} \underbrace{\overbrace{\begin{bmatrix} 1 & U_i(z) \\ 0 & 1 \end{bmatrix}}^{\text{primal lifting}}} \tag{5.26}$$

with $P_i(z)$ and $U_i(z)$ are *Laurent* polynomials and K a normalization factor. Using the perfect reconstruction property (5.6), we obtain the following factorization for the analysis

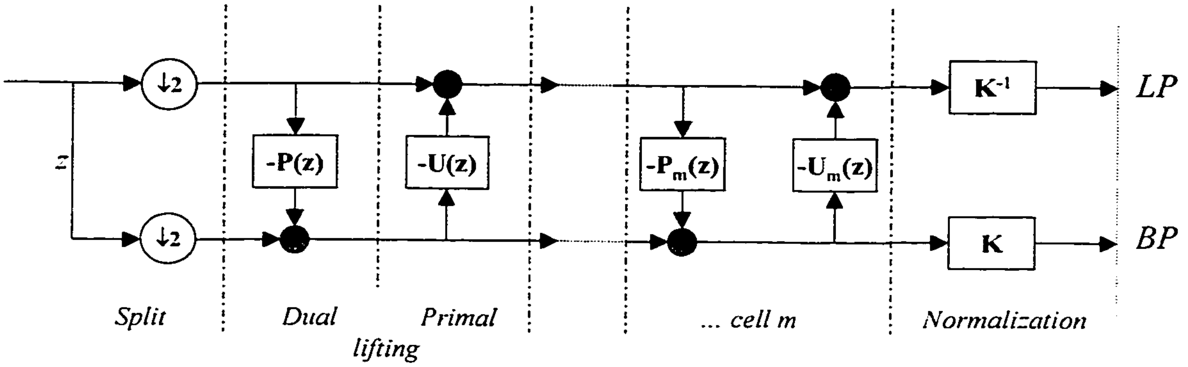


Figure 5-9: The forward lifted DWT.

polyphase matrix

$$\mathbf{H}_p(z) = \overbrace{\begin{bmatrix} 1/K & 0 \\ 0 & K \end{bmatrix}}^{\text{normalization}} \prod_{i=1}^{m} \text{predict stage} \underbrace{\overbrace{\begin{bmatrix} 1 & -P_i(z) \\ 0 & 0 \end{bmatrix}}^{\text{dual lifting}}} \text{update stage} \underbrace{\overbrace{\begin{bmatrix} 1 & 0 \\ -U_i(z) & 1 \end{bmatrix}}^{\text{primal lifting}}} \tag{5.27}$$

Figures 5-9 and 5-10 give a schematic representation of the forward and the inverse lifted DWT.

To obtain the *Laurent* polynomials $P(.)$ and $U(.)$, we can write according to the matrix relation (5.19) that

$$\begin{aligned} G_0^e(z) &= G_0^{e^{new}}(z) - G_1^e P(z) \\ G_0^o(z) &= G_0^{o^{new}}(z) - G_1^o P(z) \end{aligned} \tag{5.28}$$

with $\deg\left[G_0^{e^{new}}(z)\right] < \deg\left[G_0^e(z)\right]$ and $\deg\left[G_0^{o^{new}}(z)\right] < \deg\left[G_0^o(z)\right]$, which corresponds to

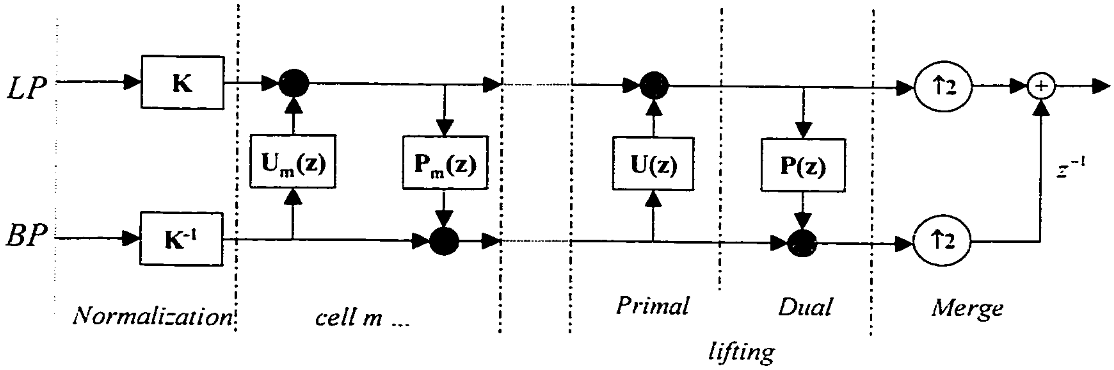


Figure 5-10: The inverse lifted DWT.

two long divisions with remainder of *Laurent* polynomials. These are performed using the *Euclidean* algorithm stated below.

Theorem 7 ***(Euclidean algorithm)*** - *Take two Laurent polynomials* $a(z)$ *and* $b(z) \neq 0$ *with* $\deg[a(z)] < \deg[b(z)]$. *Let* $a_0(z) = a(z)$ *and* $b_0(z) = b(z)$ *and iterate the following steps starting from* $i = 0$

$$\begin{aligned} a_{i+1}(z) &= b_i(z) \\ b_{i+1}(z) &= a_i(z) \bmod(b_i(z)) \\ q_{i+1}(z) &= a_i(z)/b_i(z) \end{aligned}$$

then $a_n(z) = \gcd(a(z), b(z))$ *where the number of steps is* $n \preceq \deg[b(z)] + 1.\square$

Example 8 *Let* $a(z) = a_0(z) = z^{-1} + 6 + z$ *and* $b(z) = 4 + 4z$*. The first iteration gives*

$$\begin{aligned} a_1(z) &= 4 + 4z \\ b_1(z) &= 4 \\ q_1(z) &= \frac{1}{4}z^{-1} + \frac{1}{4} \end{aligned}$$

and the next one

$$\begin{aligned} a_2(z) &= 4 \\ b_2(z) &= 0 \\ q_1(z) &= 1 + z \end{aligned}$$

where the number of steps is $n = 2$*. Thus* $a(z)$ *and* $b(z)$ *are relatively prime and we get the following factorization*

$$\begin{bmatrix} z^{-1} + 6 + z \\ 4 + 4z \end{bmatrix} = \begin{bmatrix} \frac{1}{4}z^{-1} + \frac{1}{4} & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 + z & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 4 \\ 0 \end{bmatrix}$$

Illustrative example

To summarize the procedure of factoring FIR wavelet filters into lifting steps we work out a detailed example using the biorthogonal wavelets with two vanishing moments for both the primal and dual wavelets, i.e. CDF(2,2), which has already been used to illustrate the lifting scheme in the space domain. The analysis filters are given by [20]

$$
\begin{aligned}
H_0(z) &= -\frac{1}{8}z^{-2} + \frac{1}{4}z^{-1} + \frac{3}{4} + \frac{1}{4}z - \frac{1}{8}z^2 \\
H_1(z) &= \frac{1}{4}z^{-2} - \frac{1}{2}z^{-1} + \frac{1}{4}
\end{aligned}
$$

and the analysis polyphase matrix is

$$
\mathbf{H}_p(z) = \begin{bmatrix} -\frac{1}{8}z^{-1} + \frac{3}{4} - \frac{1}{8}z & \frac{1}{4} + \frac{1}{4}z \\ \frac{1}{4}z^{-1} + \frac{1}{4} & -\frac{1}{2} \end{bmatrix}
$$

To decompose this matrix using the dual lifting, i.e.

$$
\begin{bmatrix} -\frac{1}{8}z^{-1} + \frac{3}{4} - \frac{1}{8}z & \frac{1}{4} + \frac{1}{4}z \\ \frac{1}{4}z^{-1} + \frac{1}{4} & -\frac{1}{2} \end{bmatrix} = \begin{bmatrix} H_0^{e^{new}}(z) & \frac{1}{4} + \frac{1}{4}z \\ H_1^{e^{new}}(z) & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ P(z) & 1 \end{bmatrix}
$$

we have to find the *Laurent* polynomials, $H_0^{e^{new}}(z), H_1^{e^{new}}(z)$ and the predict operator $P(z)$ such that,

$$
\begin{aligned}
-\frac{1}{8}z^{-1} + \frac{3}{4} - \frac{1}{8}z &= P(z)\left(\frac{1}{4} + \frac{1}{4}z\right) + H_0^{e^{new}}(z) \\
\frac{1}{4}z^{-1} + \frac{1}{4} &= P(z)\left(-\frac{1}{2}\right) + H_1^{e^{new}}(z)
\end{aligned}
$$

where we get by using the *Euclidean* algorithm, with $a_0 = -\frac{1}{8}z^{-1} + \frac{3}{4} - \frac{1}{8}z$ and $b_0 = \frac{1}{4} + \frac{1}{4}z$, these three different outcomes for the first equation

$$-\frac{1}{8}z^{-1}+\frac{3}{4}-\frac{1}{8}z = \left\{\begin{array}{l} \left(-\frac{1}{2}z^{-1}+\frac{7}{2}\right)\left(\frac{1}{4}+\frac{1}{4}z\right)-z \\ \left(-\frac{1}{2}z^{-1}-\frac{1}{2}\right)\left(\frac{1}{4}+\frac{1}{4}z\right)+1 \\ \left(\frac{7}{2}z^{-1}-\frac{1}{2}\right)\left(\frac{1}{4}+\frac{1}{4}z\right)-z^{-1} \end{array}\right\}$$

This fact is due to the non-uniqueness of *Laurent* polynomial division. It is not however perceived as a flaw of the factorization technique but rather as an advantage because by selecting the second outcome of the factorization we are able to maintain symmetry. This unravel another facet of the flexibility of the lifting scheme factorization. This choice is also suitable with the factorization of the second equation as

$$\frac{1}{4}z^{-1}+\frac{1}{4} = \left(-\frac{1}{2}z^{-1}-\frac{1}{2}\right)\left(-\frac{1}{2}\right)+0$$

The polyphase matrix is then factored using a dual lifting as

$$\mathbf{H}_p(z) = \begin{bmatrix} 1 & \frac{1}{4}+\frac{1}{4}z \\ 0 & -\frac{1}{2} \end{bmatrix} \overbrace{\begin{bmatrix} 1 & 0 \\ -\frac{1}{2}z^{-1}-\frac{1}{2} & 1 \end{bmatrix}}^{\text{dual lifting}}$$

The next step consists then in extracting the *Laurent* polynomials, $H_0^{o^{new}}(z), H_1^{o^{new}}(z)$

and $U(z)$ as,

$$\mathbf{H}_p(z) = \begin{bmatrix} 1 & H_0^{o^{new}}(z) \\ 0 & H_1^{o^{new}}(z) \end{bmatrix} \begin{bmatrix} 1 & U(z) \\ 0 & 1 \end{bmatrix} \overbrace{\begin{bmatrix} 1 & 0 \\ -\frac{1}{2}z^{-1} - \frac{1}{2} & 1 \end{bmatrix}}^{\text{dual lifting}}$$

which amounts to solve the following equations,

$$\begin{aligned} \frac{1}{4} + \frac{1}{4}z &= U(z) + H_0^{o^{new}}(z) \\ -\frac{1}{2} &= 0 + H_1^{o^{new}}(z) \end{aligned}$$

A quick inspection gives the primal lifting and consequently leads to the complete factorization of the polyphase matrix, i.e.

$$\mathbf{H}_p(z) = \overbrace{\begin{bmatrix} 1 & 0 \\ 0 & -\frac{1}{2} \end{bmatrix}}^{\text{normalization}} \overbrace{\begin{bmatrix} 1 & \frac{1}{4} + \frac{1}{4}z \\ 0 & 1 \end{bmatrix}}^{\text{primal lifting}} \overbrace{\begin{bmatrix} 1 & 0 \\ -\frac{1}{2}z^{-1} - \frac{1}{2} & 1 \end{bmatrix}}^{\text{dual lifting}}$$

Using the predict and update operators coefficients, the lifted (2,2) biorthogonal wavelet transform can be implemented completely in the space domain using the wiring of figure 5-2.

5.3.3 Orthogonal and biorthogonal wavelet filters factorization

To complete this section, the versatility of the lifting scheme in performing the factorization of any FIR wavelet transform will be demonstrated. The first example is the *Daubechies* orthogonal DWT with two vanishing moments and the second one is the biorthogonal (9,7) wavelet transform.

Daubechies orthogonal wavelet filter

The filters are given by

$$\begin{aligned} H_o(z) &= c_0 + c_1^{-1} + c_2 z^{-2} + c_3 z^{-3} \\ H_1(z) &= -c_3 z^2 + c_2 z - c_1 + c_0 z^{-1} \end{aligned}$$

with

$$\begin{array}{ll} c_o = \frac{1+\sqrt{3}}{4\sqrt{2}} & ; c_1 = \frac{3+\sqrt{3}}{4\sqrt{2}} \\ c_2 = \frac{3-\sqrt{3}}{4\sqrt{2}} & ; c_1 = \frac{1-\sqrt{3}}{4\sqrt{2}} \end{array}$$

The polyphase matrix,

$$\mathbf{H}_p(z) = \begin{bmatrix} c_o + c_2 z^{-1} & c_1 + c_3 z^{-1} \\ -c_3 z - c_1 & c_2 z + c_0 \end{bmatrix}$$

is then factored into two dual lifting, one primal lifting and a normalization, as [22]

$$\mathbf{H}_p(z) = \overset{\text{normalization}}{\begin{bmatrix} \frac{\sqrt{3}+1}{\sqrt{2}} & 0 \\ 0 & \frac{\sqrt{3}-1}{\sqrt{2}} \end{bmatrix}} \overset{\text{dual lifting 2}}{\begin{bmatrix} 1 & 0 \\ z & 1 \end{bmatrix}} \overset{\text{primal lifting}}{\begin{bmatrix} 1 & \frac{\sqrt{3}}{4} + \frac{\sqrt{3}-2}{4} z^{-1} \\ 0 & 0 \end{bmatrix}} \overset{\text{dual lifting 1}}{\begin{bmatrix} 1 & 0 \\ -\sqrt{3} & 1 \end{bmatrix}}$$

The DWT can thus be implemented in the space domain using the following procedure:

$$\begin{aligned}
d_k^1 &\longleftarrow a_{2k+1}^0 - \sqrt{3}a_{2k}^0 \\
a_k^1 &\longleftarrow a_{2k}^0 + \frac{\sqrt{3}}{4}d_k^1 + \frac{\sqrt{3}-2}{4}d_{k+1}^1 \\
d_k^2 &\longleftarrow d_k^1 + a_{k-1}^1 \\
a_k &\longleftarrow \frac{\sqrt{3}+1}{\sqrt{2}}a_k^1 \\
d_k &\longleftarrow \frac{\sqrt{3}-1}{\sqrt{2}}d_k^2
\end{aligned}$$

The inverse transform is straightforward and is obtained by reversing the above operations and flipping the signs, i.e.

$$\begin{aligned}
d_k^2 &\longleftarrow \frac{\sqrt{3}+1}{\sqrt{2}}d_k \\
a_k^1 &\longleftarrow \frac{\sqrt{3}-1}{\sqrt{2}}a_k \\
d_k^1 &\longleftarrow d_k^2 - a_{k-1}^1 \\
a_{2k}^0 &\longleftarrow a_k^1 - \frac{\sqrt{3}}{4}d_k^1 - \frac{\sqrt{3}-2}{4}d_{k+1}^1 \\
a_{2k+1}^0 &\longleftarrow d_k + \sqrt{3}a_{2k}^0
\end{aligned}$$

FBI biorthogonal filters

Here we consider again the famous (9-7) biorthogonal filter known as the FBI filter pair (see section 3.3.3). The analysis lowpass filter H_0 has 9 coefficients while the synthesis lowpass filter G_0 has 7 coefficients. Both highpass filters H_1 and G_1 have 4 vanishing moments. Applying the *Euclidean* algorithm to the analysis filter, i.e. the longer one, the DWT is factored into two dual stages and two primal stages as [22],

$$\mathbf{H}_p(z) = \underbrace{\begin{bmatrix} K & 0 \\ 0 & K^{-1} \end{bmatrix}}_{\text{normalization}} \underbrace{\begin{bmatrix} 1 & 0 \\ \gamma\left(1+z^{-1}\right) & 1 \end{bmatrix}}_{\text{dual lifting 1}} \underbrace{\begin{bmatrix} 1 & \delta\left(1+z^{-1}\right) \\ 0 & 0 \end{bmatrix}}_{\text{primal lifting 1}}$$

$$\underbrace{\begin{bmatrix} 1 & 0 \\ \alpha\left(1+z^{-1}\right) & 1 \end{bmatrix}}_{\text{dual lifting 2}} \underbrace{\begin{bmatrix} 1 & \beta\left(1+z^{-1}\right) \\ 0 & 0 \end{bmatrix}}_{\text{primal lifting 2}}$$

with

$$\alpha = -1.586 \quad ; \beta = -0.052$$
$$\gamma = 0.882 \quad ; \delta = 0.443$$

The factorization, which has the additional advantage of being symmetric, is implemented as follows

$$d_k^1 \leftarrow a_{2k+1}^0 + \alpha(a_{2k}^0 + a_{2k+2}^0)$$
$$a_k^1 \leftarrow a_{2k}^0 + \beta(d_k^1 + d_{k-1}^1)$$
$$d_k^2 \leftarrow d_k^1 + \gamma(a_k^1 + a_{k+1}^1)$$
$$a_k^2 \leftarrow a_k^1 + \delta(d_k^2 - d_{k-1}^2)$$
$$a_k \leftarrow K a_k^2$$
$$d_k \leftarrow \frac{1}{K} d_k^2$$

The time-domain implementation of the (9,7) biorthogonal DWT for a 8-sample input

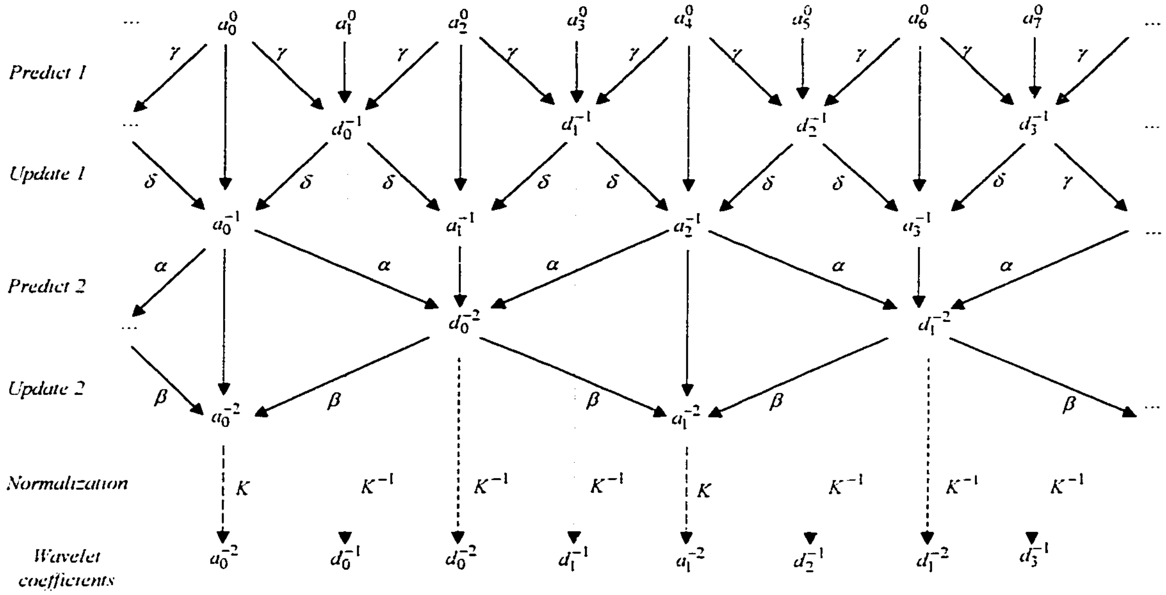


Figure 5-11: Time-domain implementation of the lifted (9,7)-biorthoghonal DWT.

signal is depicted in figure 5-11 that clearly exhibits a symmetric pattern.

Chapter 6

WAVELET SHRINKAGE AND MDL CODING

This chapter covers the last stage of the proposed seismic compression scheme. It treats the challenging problem of wavelet coefficients encoding. New statistical encoding strategies are used. These are based on two efficient and powerful concepts, namely the *wavelet shrinkage principle* developed by *Donoho* and co-workers [28],[29],[30], and the *minimum description length* (MDL) developed by *Rissanen* [79]. The proposed encoding engine is an adaptive version of the *threshold USQ* (TUSQ) [104], [107]. Indeed, instead of selecting the zero bin in an *ad hoc* manner as twice the bin size [76], an adaptive wavelet shrinkage approach is adopted to determine the best zero-bin based on the statistics of each detail subband population. The design of the TUSQ outside the dead-zone consists of finding the number of bins m and their size Δ, based on results from *Kolmogorov* theory [64] rather than the conventional *Shannon* theory [50]. The *Kolmogorov* theory is complex but its practical version, i.e. the MDL, is more tractable yet very efficient. The rationale of the MDL principle is to select a model with low complexity, i.e. having a simple representation or a reasonable number of parameters, and fitting well the quantized coefficients. Finally, the binary coding stage is lossless and consists of a *run-length coding* (RLC) followed by a *Huffman* coder. It is similar to the one prevailing in the JPEG2000 standard [52].

6.1 Wavelet Shrinkage

6.1.1 Heuristic

Formally, the wavelet shrinkage has been proposed by *Donoho* and *Johnstone* [28],[30] for near-optimal non-parametric regression. The heuristic of this approach is that for spatially inhomogeneous signals (e.g. seismic signals) most of the information is concentrated in a small subset of the wavelet space. This is mainly due to the inherent space and frequency localization properties of the wavelet transform. Elsewhere, under the *additive white Gaussian noise* (AWGN) assumption[1], noise contaminates all the coefficients equally across all the scales, whereas only a very few wavelet coefficients bear the characteristic features of the signal. Wavelet shrinkage exploits those spectral and structural differences of the underlying signal and noise across the scales to efficiently separate their respective components. In general, wavelet shrinkage could be performed using two thresholding operators, namely the hard- and the soft-thresholding operators. However, its diversity resides in the plethora of techniques available to determine the threshold value(s) based

---

[1] which is still the best choice in the case the true nature of the noise is unknown.

on the statistics of the wavelet coefficients.

6.1.2 Thresholding operators

An intuitive way for achieving thresholding consists of applying the following non-linear hard-thresholding operator,

$$\eta_\lambda(d_{jk}) \equiv d_{jk}.\mathbb{I}\{|d_{jk}| > \lambda\} \tag{6.1}$$

which can be seen as a keep-or-kill rule as it keeps the wavelet coefficients whose absolute value is larger than the threshold λ and kills all the rest.

In practice hard-thresholding may create spots in the reconstructed data especially when the noise energy is significant. Shrinking thus the value of the survived coefficients to zero by an amount equals to the threshold value (see figure 6-1) using the following *soft-thresholding* operator,

$$\eta_\lambda(x) \equiv sgn(x).\max(|x| - \lambda, 0) \tag{6.2}$$

is in general preferable because of various advantages. From a visual point of view, the reconstructed data offer a more pleasant aspect and do not exhibit visible artifacts. Recall that some artifacts can be cured by selecting more regular reconstruction wavelet filters. Moreover, soft-thresolding uses a continuous function (6.2), which is computational more tractable than the discontinuous function (6.1), for the determination of the threshold values.

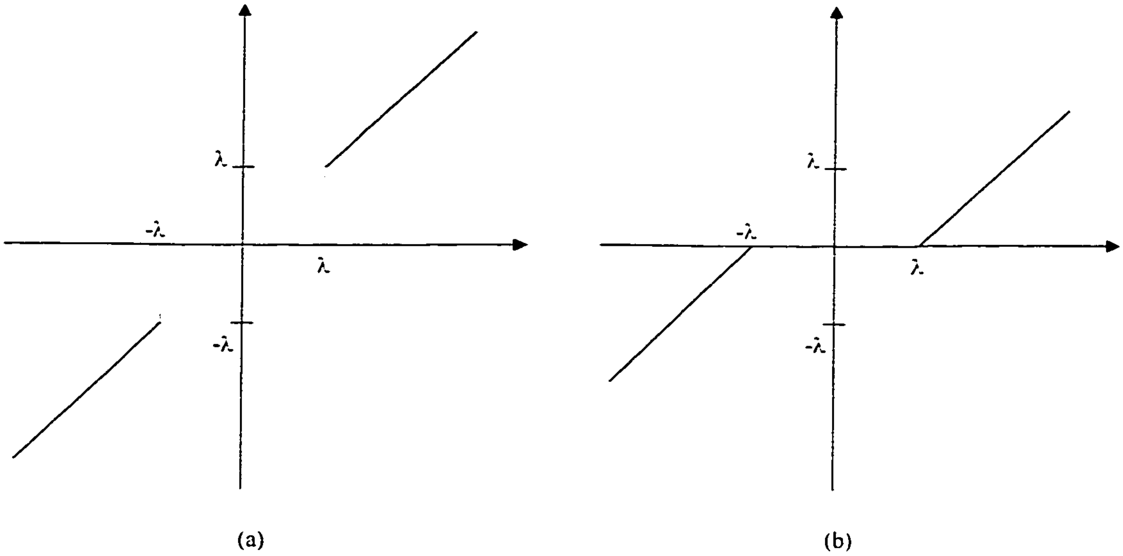


Figure 6-1: Schematic representation of the thresholding operators. (a) Hard-thresholding, (b) Soft-thresholding.

The selection of the threshold value(s) is a very delicate and important statistical problem[2]. On one hand, killing too many wavelet coefficients may lead to an important *bias* in the reconstructed data. On the other hand, small thresholds lead to a poor compression gain. Thus, threshold(s) selection should strike the balance between closeness to fit between the original and the reconstructed data and the degree of sparsity of the wavelet coefficients. Note that the bias introduced by the soft-thresholding estimator is less than the one introduced by even the optimal hard-thresholding estimator [2]. In figures 6-1-(a) and -(b), we give a schematic representation of the operators (6.1) and (6.2).

---

[2] In this context, the thresholding operators are called estimators.

6.1.3 Wavelet shrinkage adaptivity

Spatial adaptivity

Donoho and *Johnston* [30] showed that the wavelet shrinkage operator is endowed with a built-in spatial adaptivity. This feature fits very well seismic data, which consist of relatively flat portions, tolerating thus a greater amount of smoothing, and some regions with one or more important features, allowing then less smoothing. The wavelet shrinkage estimator is able to discern from the wavelet coefficients solely where more smoothing is needed, and then to apply the needed amount of thresholding to generate data compression, denoising or both.

Scale and orientation adaptivity

The DWT generates wavelet coefficients lying in subbands that consist of a residual image and three images emphasizing the details in the horizontal, vertical and diagonal orientations. The scale-adaptivity is obtained when we derive a specific threshold for each detail subband at each scale. On the other hand, the orientation adaptivity is obtained by exploiting the anisotropy of the details subbands. Note that in practice we do not threshold and even not quantize the approximation wavelet coefficients at the coarsest scale because these are measuring the gross structure of the data. Furthermore, the corresponding approximation subband or residual image has a relatively small size and does not contribute significantly to the coding gain. In view of guaranteeing the best shrinkage operators adaptivity, two different techniques are proposed for the threshold selections, namely the SURE principle [28],[30] and the *Birgé-Massart* strategy [6].

For the sake of comparison, a single and global threshold for all the details coefficients regardless of their scale and orientations is also considered. The thresholding rule is known as *VisuShrink* and the *universal* threshold is given by [31],

$$\lambda = \sqrt{2 \log n} \tag{6.3}$$

where n is the number of wavelet detail coefficients.

6.2 Adaptive Threshold Selection Rules

6.2.1 SureShrink operator

Stein unbiased risk estimate (SURE)

SURE has been initiated by *Stein* for mean estimation of a multivariate normal distribution [88] and has been successfully applied for function smoothing by *Donoho* et al. [28],[29],[30]. The foundation of the SURE principle is based on the fact that for nearly arbitrary nonlinear biased estimator, the loss or risk can be estimated unbiasedly. First, the SURE principle will be outlined for the general setting of nonlinear estimation, then it will be shown how the SURE leads to the *SureShrink* [30], the wavelet shrinkage operator.

Consider an empirical data vector y of dimension N given by

$$y_i = f_i + e_i \quad i = 0, ..., N-1 \tag{6.4}$$

where f_i are samples of the deterministic signal f and e is *Gaussian* white noise with *independent identical distribution* (i.i.d) $\mathcal{N}(0, \sigma)$. The objective is to find the best estimate of the function f in the mean square sense by minimizing the *mean square error* (MSE) risk defined as,

$$R(\check{f}, f) = \frac{1}{N} \left\| \check{f} - f \right\| = \frac{1}{N} \sum_{i=0}^{N-1} (\check{f}_i - f_i)^2 \tag{6.5}$$

In practice the MSE (6.5) can never be computed exactly because it relies on the unknown exact value of the signal f. However, the MSE can be estimated based on the SURE principle. Indeed, if we consider the following estimate for the unknown signal f,

$$\check{f}(\zeta) = \zeta + \eta(\zeta) \tag{6.6}$$

where $\eta(\zeta)$ is a weakly differentiable function from R^N to R^N, then an unbiased estimator for the MSE risk is the SURE defined as [88],

$$R^{SURE}(\check{f}(\zeta), f) = N + E\{\|\eta(\zeta)\|^2 + 2\nabla_\zeta \eta(\zeta)\} \tag{6.7}$$

where ∇ is the vector differential operator of first partial derivatives, i.e.

$$\nabla_\zeta \eta(\zeta) \equiv \sum_{i=0}^{N-1} \frac{\partial}{\partial \zeta_i} \eta_i \tag{6.8}$$

SureShrink operator

The scale and orientation-dependent threshold λ^j is derived by considering the following soft-thresholding function,

$$\eta_\lambda(d^j) = sgn(d_k^j)\left(\left|d_k^j\right| - \lambda^j\right) \# \left(\left|d_k^j\right| > \lambda^j\right) \tag{6.9}$$

in lieu of the weakly differentiable function defined above. The SURE for the soft-threshold estimator is then given by [30],

$$R_\lambda^{SURE} = 2^j - 2.\# \left\{ \left|d_k^j\right| \preceq \lambda^j\} + \sum_{i=0}^{2^j-1} \min\{\left|d_k^j\right| \preceq \lambda^j \right\}^2 \tag{6.10}$$

and the scale and orientation-dependent threshold stems from the minimization of the SURE, i.e.

$$\lambda^j = \arg\min_{\lambda \succeq 0} R_\lambda^{SURE}(\lambda^j, d_k^j) \tag{6.11}$$

The resulting soft-thresholding function $\eta_\lambda(d_k^j)$ is referred to as *SureShrink* operator.

6.2.2 Birgé-Massarat operator

The *Birgé-Massart* strategy is based on new results on adaptive functional estimation in regression context [6]. The authors used the methods of penalization to develop new adaptive estimators in the context of wavelet expansions in non-homogeneous *Besov* spaces. Theoretically, the problem of threshold selection can be considered as an approximation problem. From a practical point of view, this sophisticated strategy has a very clean implementation [71]. At the decomposition level $J_0 + 1$ all the wavelet coefficients are kept. Then for each level or scale j from 1 to J_0, only the wavelet coefficients greater than the following scale-dependent threshold,

$$\lambda^j = \frac{m}{(J_0 + 1 - j)^\alpha} \tag{6.12}$$

are kept.

The parameters J_0, m, and α, which define the strategy are respectively the decomposition depth that fixes the low frequency cutoff, the length of the coarsest approximation coefficients, i.e. at level $J_0 + 1$ and an universal parameter equals to 1.5 for compression and 3 for denoising [6].

6.2.3 Compression versus denoising

Wavelet shrinkage can also be used for denoising tasks [31]. However, it has been shown that compression using wavelet shrinkage has a denoising effect [65]. This is mainly due to the dead-zone of the TUSQ, which kills the negligeable wavelet coefficients as these are most likely due to incoherent noise.

This contribution is not only a continuation on the theme of simultaneous compression and denoising [84] using wavelet shrinkage, the challenging problem of removing the ground-roll [1] is also dealt with.

6.3 Minimum Description Length (MDL) Encoding

The second stage of the adaptive quantizer concerns the quantization of the survived detail coefficients. In theory, the objective of compression is to keep the distortion low, while achieving a high coding gain. For a given bit budget, we wish to allocate the least number of bits to code the description of the source, which may be the thresholded details wavelet coefficients. These requirements are conflicting because in order to minimize the distortion, a sophisticated model of the source is required, which in turn requires the coding of its various parameters. To solve this problem a simple and elegant solution based on the MDL formalism is proposed.

6.3.1 Kolmogorov complexity

MDL originated from the proposition of *Solomonoff* [79] to use a *universal computer language* to describe a given sequence of observations. Then *Kolmogorov* [64] proposed to quantify the information present in the latter by considering its complexity[3]. Formally, we can define the *Kolmogorov complexity* as the length of the shortest computer program that generates a sequence of observations using a *universal turning machine.* However, the drawback of the *Kolmogorov* complexity is its lack of computational tractability. The MDL principle proposed by *Rissanen* [79] can be regarded as a practical version of the *Kolmogorov* complexity. Indeed, the MDL considers only description methods that allow the computation of the length of the shortest description of a sequence attainable using

---

[3] *Shannon* complexity uses the entropy as a measure of the information present in a given sequence of observations.

those description methods. There are many variants of the MDL principle [64],[79], but the *two-part code* MDL is conceptually suitable for the design of the MDL-based adaptive quantizer.

6.3.2 Two-part code MDL

The rationale of the MDL principle is to select a model with low complexity, i.e. having a simple representation or a reasonable number of parameters, and fitting well the subband detail wavelet coefficients. Thus, the MDL principle is based on striking the balance between regularity and randomness in the data to compress. In other terms, it provides a trade-off between *model complexity* and *goodness-of-fit* on the data. In this context, the best model is always the one with minimum complexity and that optimally uses regularity to reconstruct the data.

Formally, the MDL principle stipulates that given a sequence of data and an effective enumeration of the appropriate alternative models to explain the data, the best model is the one that minimizes the sum of:

1. the length, in bits, of the description of the model; and
2. the length, in bits, of the description of the data when encoded with that model.

Mathematically, the two-part code MDL is expressed as,

$$\mathcal{M}^{mdl} = \arg \min_{\mathcal{M} \in \mathfrak{M}} \{\mathcal{L}_{C_2}(d_\lambda^j | \mathcal{M}) + \mathcal{L}_{C_1}(\mathcal{M})\} \tag{6.13}$$

where $\mathcal{M}$ is a model belonging to some class of models $\mathfrak{M}$, d_λ^j a given sequence of observations (e.g., the thresholded wavelet coefficients), $\mathcal{L}_{C_1}$ the description length for encoding the model $\mathcal{M}$ using the code C_1, and $\mathcal{L}_{C_2}$ the description length for encoding the wavelet coefficients d^j_λ given the model $\mathcal{M}$ and using the code C_2.

To give an intuition about the MDL principle, let us consider some typical scenarios. To minimize the distortion between a signal and its description, we can choose the signal itself as a model, in which case there is no distortion but at the cost of using many bits to encode the signal. Thus no compression occurs. At the other frontier, we can use the zero function, which requires essentially zero bit to encode at the expense of a high distortion. In the middle ground, a parametric model may yield a total description length as the number of bits needed to encode the data given the model, plus the number of bits needed to encode the number of parameters and the parameters value. The MDL helps in selecting among a given class of models the one achieving a good compromise between fitting the signal well and having a reasonable number of parameters. In the case of the wavelet quantization there is no deterministic model available. The transform coefficients consists of many small coefficients and a relatively few large coefficients. Consequently, they form a histogram with a peak at zero and symmetric about zero and can be fairly modeled as distributions. Thus a connection between codes and probability distributions is of order. An elegant illustration of such a connection is given by the *Shannon* code, which assigns a high probability to more frequently occurring symbols (i.e. short symbols). In the same spirit, the MDL principle can be considered as a *maximum probability* principle, since short code lengths correspond to high probabilities and vice-versa. However a noticeable difference between the MDL and the *maximum likelihood (ML)* estimator resides in the regularization penalty that the MDL uses to counterbalance the effect of the data size [65].

For a probabilistic model with $\mathcal{K}$ fixed parameters and given a sequence of observations of length N, the *MDL* principle tells us to pick the following model [84],

$$\mathcal{M}^{mdl} = \arg\min_{\mathcal{M}\in\mathfrak{M}} \{-\log_2 P(d^j_\lambda|\mathcal{M}) + \mathcal{L}_{C_1}(\mathcal{M}) + \frac{\mathcal{K}}{2}\log_2 N\} \tag{6.14}$$

where the last term represents the penalty to code the decimal part of the parameters when these are truncated up to a precision δ [79]. This term can be ignored in the minimization and will be dropped. Elsewhere, when we look at the MDL in a non-probabilistic framework (e.g. polynomial fitting), it turns out that a good measure for the trade-off between complexity and goodness-of-fit is the *squared error function*. Thus the probability distribution P, which achieves the best trade-off, is simply the n-fold product distribution of a conditional *Gaussian* pdf[4]. Next, we will show how to determine $\mathcal{L}_{C_1}$ in order to design the USQ, which will be applied to the wavelet coefficients outside the dead-zone(s).

6.3.3 MDL-based quantization

Consider an arbitrary detail subband of size $2^j x 2^j$ at a given scale j and orientation, whose population d^j_λ have been thresholded using either the *sureshrink* operator or the *Birgé-Massart* strategy. It has been shown in [68],[104] that the threshoded wavelet coefficients

---

[4] Even if the data is generated according to a completly different distribution, it may still be useful and harmless to model it using a normal distribution.

can be fairly modeled with a *generalized Gaussian distribution* (GGD), i.e.

$$GG_{\alpha,\beta}(x) = C(\alpha, \beta)e^{-(\alpha|x|)^{\beta}} \tag{6.15}$$

where $C(\alpha, \beta) = \frac{\alpha\beta}{2\Gamma(\frac{1}{\beta})}$ and $\Gamma(t)$ is the gamma function. However, the *Laplacian* distribution ($\beta = 1$) is particularly interesting because it is frequently used to make the analysis more tractable [107]. The design of the *entropy-constrained scalar quantizer (ECSQ)* for a *Laplacian pdf* has been dealt with by *Sullivan* [93]. However, from a practical point of view, a simple but yet efficient approximation of *ECSQ* is the *TUSQ*. As the wavelet coefficients fed to the quantizer are already thresholded the *USQ* fairly models the set of quantized wavelet coefficients $\bar{d}^j_\lambda$. The *USQ* is characterized by $2m + 1$ reconstruction levels of equal bin size Δ indexed as $l = -m, ..., -1, 0, 1, ..., m$ [50]. Consequently, the code C_1 reduces to the *Shannon*-code and the description length $\mathcal{L}_{C_1}$ becomes,

$$\mathcal{L}_{C_1}(\mathcal{M}) = \mathcal{L}_{C_1}(m, \Delta) + \mathcal{L}_{C_1}(\bar{d}^j_\lambda | m, \Delta) \tag{6.16}$$

where the second term is the length to code the number of bins m and the bin size Δ.

Since the quantized wavelet coefficients can be modeled as,

$$\bar{d}^j_\lambda = d^j_\lambda + \epsilon \tag{6.17}$$

where quantization noise ϵ is i.i.d $\mathcal{N}(0,\sigma^2)$, this implies that $\bar{d}^j_\lambda \sim \mathcal{N}(0,\sigma^2)$. Consequently

$$\mathcal{L}_{C_1}(m,\Delta) = \frac{1}{2\ln 2\sigma^2}\sum_{i=1}^{2^j}(d^j_i - \bar{d}^j_i)^2 \tag{6.18}$$

On average, the smallest number of bins needed to encode the bin indices given the model $\mathcal{M}$ is [14],[93],

$$\mathcal{L}_{C_1}(\bar{d}^j_\lambda|m,\Delta) = -\sum_{l=-m}^{m} K_l \log_2 \frac{K_l}{2^{2j}} \tag{6.19}$$

Finally, the following *MDL* minimization process,

$$(\bar{d}^j_\lambda, m, \Delta)^{MDL} = \arg\min\{\mathcal{L}_{C_1}(\mathcal{M})\} \tag{6.20}$$

where K_l is the number of the wavelet coefficients belonging the bin l, yields the best model of the USQ in the sense of *Kolmogorov.*

6.3.4 Estimation of the noise variance

From equation (6.18), the noise variance σ^2 is needed. In general the latter is unknown but could be estimated from the wavelet coefficients at the finest scale in the diagonal direction. The rationale of this approach is that in general, the top level coefficients consist in mostly of noise. However, since there might be some signal features present at the finest scale, it is safe to use the robust median absolute deviation (MAD) estimator, defined as [29]

$$\breve{\sigma} = \frac{MAD(|d_1^D|)}{0.6745} \tag{6.21}$$

where $\breve{\sigma}$ is the estimated standard deviation.

Chapter 7

RESULTS AND DISCUSSION

The proposed compression strategies are applied to a real seismic data set provided by *Saudi ARAMCO*. The data set consists of a time-migrated section of the *Midyan Basin* in the *Red Sea* (KSA) [73]. First, the problem of borders corrections is dealt with using both orthogonal and biorthogonal wavelet systems. Then, the ability of the orthogonality, the symmetry, the smoothness and the support size of the 2-D wavelet bases to compact the energy and to reduce the entropy of the input data is evaluated. Visual as well as statistical tools, such as the correlation functions and the histograms, are used. After experimentally selecting the most suitable wavelet bases for transforming the seismic data, the problem of wavelet shrinkage is tackled. As the design of the proposed adaptive quantizer is based on the statistics of the detail subband coefficients, the distribution of the coefficients is first analyzed using histogram fitting technique and quantile-to-quantile plots. Then an adaptive dead-zone for the TUSQ is designed for every detail subband using the SURE principle [30] and the *Birgé-Massarat* strategy [6] respectively. Finally, in view of demonstrating the ability of wavelet shrinkage to filter out seismic coherent noise, the challenging problem of ground-roll removal is considered using a different dataset described in the reference manual of *Yilmaz* [109] and available from the *Colorado School of Mines* anonymous ftp site at ftp.cwp.mines.edu.

7.1 Midyan Data Set

In *Saudi Arabia*, most of the strata are gently dipping at angles less than a few degrees. However, along the *Red Sea* area, the subsurface structures become more complex due to *Miocene* rifting and to salt diapirism. The *Midyan Basin* is located at the junction between the *Red Sea* and the *Gulf of Aqaba* rifts. This coastal basin is highly tectonized by conjugated faults (N35&N130). The fault throw can reach several thousand meters around basement highs surrounded by major half graben troughs. Within the pre-salt sequence, main structural elements are large extensional blocks including pre- and syn-rift tilted and faulted strata. Above the salt, a combination of extensional (growth fault with associated roll-over) and halo-kinetic tectonics dominates the post-salt sequence [73]. These complex structural features appear on the interpreted *Midyan* time-migrated section of figure 7-1, which consists of 2838 CMP of 626 samples. In figure 7-2, the seismic image used for simulation purposes is shown using a variable density display (VDD) mode (see section 1.2.2). By using the variable area display (VAD) mode of figures 7-3 and 7-4, it appears that the seismic image of figure 7-2 consists of waveforms rather than pixels.

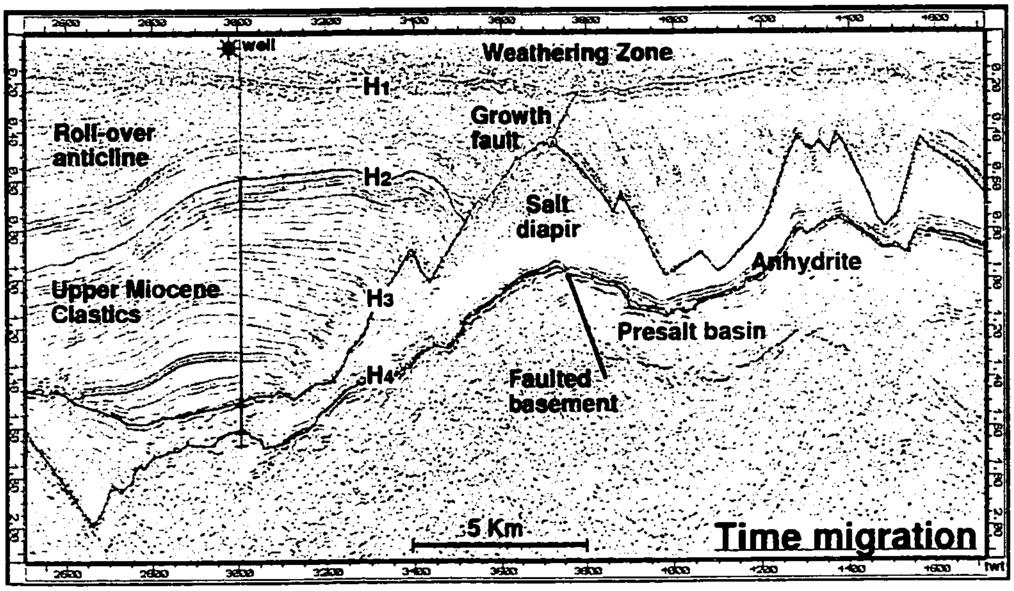


Figure 7-1: Interpreted *Midyan* time-migated seismic section (Courtesy [73])

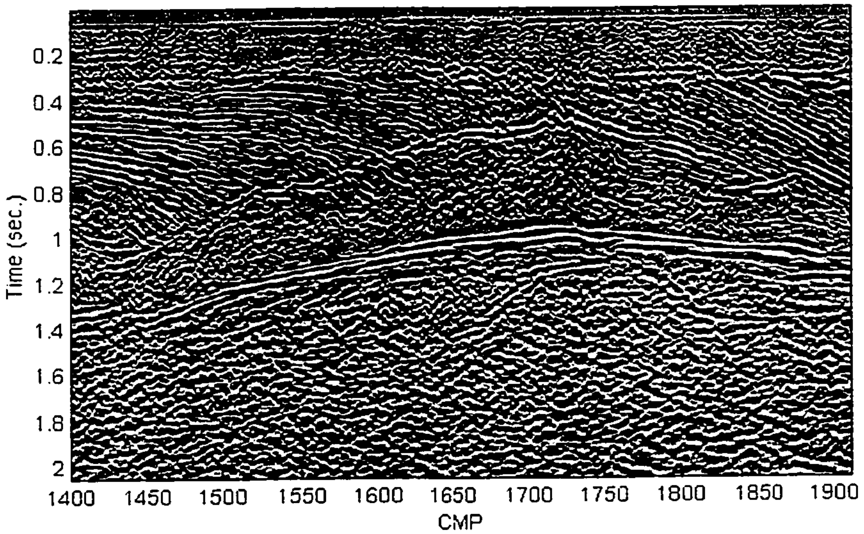


Figure 7-2: Excerpt of the original *Midyan* time-migrated section.

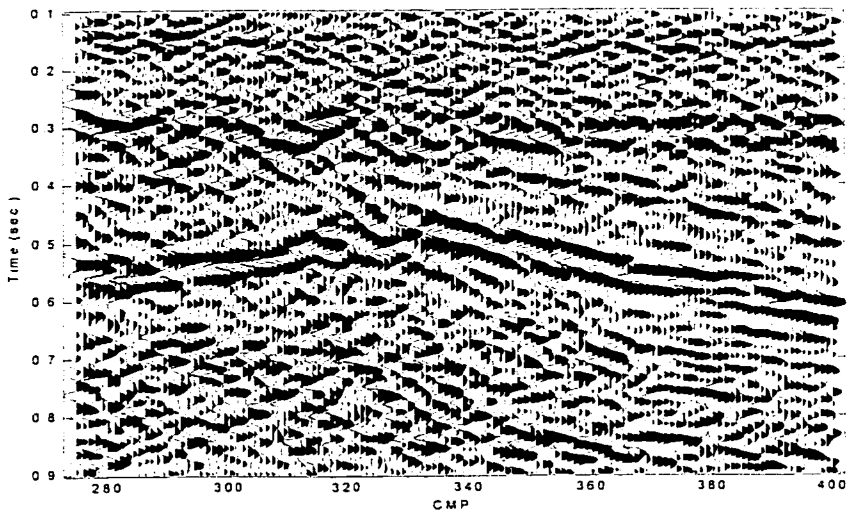


Figure 7-3: Variable area plot of an excerpt of the *Midyan* section focusing the salt diapirism.

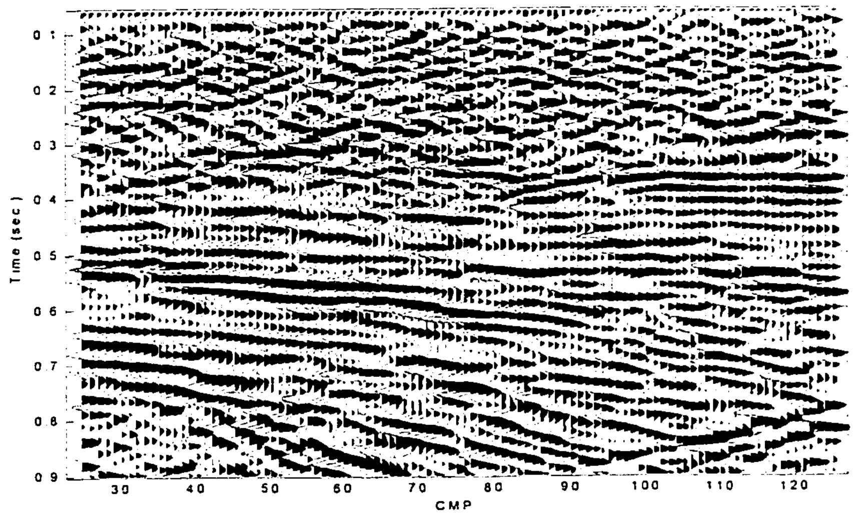


Figure 7-4: Variable area plot of an excerpt of the *Midyan* section focusing the clastics and the weathering zone.

7.2 Visual and Statistical Wavelet Analysis

The following wavelet analysis study concerns the borders corrections issue and the comparison of the performance of the 2-D wavelet filters in terms of the frequency selectivity, the entropy and the energy of the wavelet coefficients. The objective of this preliminary analysis is twofold. Firstly, it aims at verifying some theoretical assumptions related to the wavelets properties and at shedding the light on the mechanisms of the seismic wavelet expansion. Secondly, it aims at experimentally selecting the most suitable 2-D wavelet bases for subsequent compression. Therefore, the subsequent analysis exploits:

- The 2-D wavelet system bases and their spectra
- The wavelet coefficients
- The waveforms of reconstructed subbands

7.2.1 Borders corrections

The performance of the three borders correction techniques described in section 4.3 are experimentally assessed . In figure 7-5, the effect of the zero padding, the periodic and the symmetric extensions are compared using the residual subband at the first level of the wavelet expansion obtained with the most-asymmetric orthogonal wavelet system with $p = 9$ vanishing moments. The zero padding technique pads the wavelet coefficients at the beginning and the end of each seismic trace with zeros. This results in $l = \left\lceil \frac{2p-1}{2} \right\rceil - 1$ zero traces in the horizontal direction. The effect of zero padding in the vertical direction is not perceivable as the original traces are already zeroed above the weathered zone.

However, because of the many zeros introduced in the data, the numerically stability of the wavelet transform deteriorates quickly as the wavelet filters lengths increase. The periodic extension correction , which assumes that the seismic data is periodic yields boundary artifacts in both the vertical and horizontal directions as shown in figure 7-5. Therefore, the technique adopted in the rest of this experimental study is the symmetric extension correction. It is also called reflection[1] method because the original data samples are reflected at the boundaries by duplicating both of the endpoint samples and then extended periodically. In addition of being numerically stable and fast, the symmetric extension ensures perfect reconstruction for symmetric biorthogonal wavelet filters. In figure 7-5, it can be verified that the symmetric extension does not exhibit noticeable artifacts besides offering a visually pleasant aspect. Furthermore, the effect of the filters lengths on the performance of the symmetric extension by using respectively short, medium, and long biorthogonal wavelet filters is investigated. As it can be inferred from figure 7-6, the effect of the filters lengths is not noticeable because both the correction technique and the filters are symmetric. However, this is not the case for the periodic extension where the importance of the artifacts in both directions is proportional to the filters length as illustrated by figure 7-7.

7.2.2 Bases performances

The wavelet transform decorrelates the input data by reducing its entropy and confining most of its energy in a small number of wavelet coefficients. The performance of the

---

[1] To be distinguished from the seismic reflection method (see chapter 1).

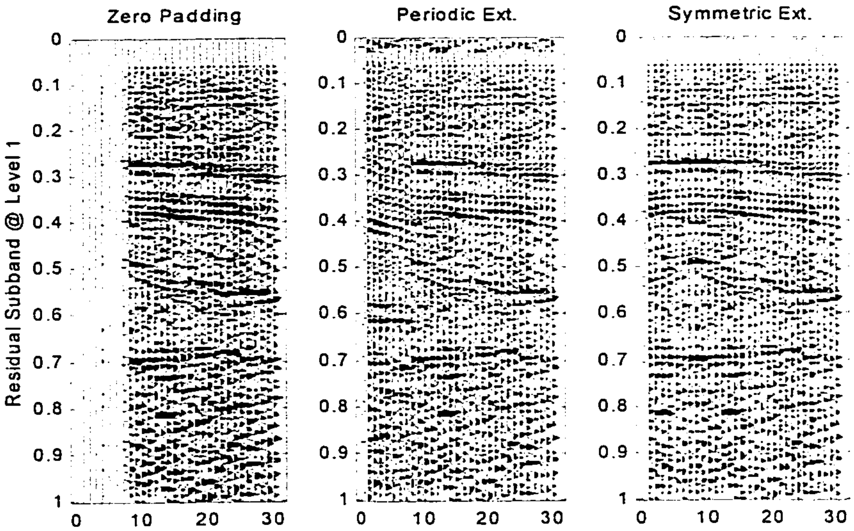

Zero Padding
Periodic Ext.
Symmetric Ext.

Figure 7-5: Performance of the zero padding, periodic and symmetric extensions using the most-asymmetric orthogonal wavelet system with $p = 9$ vanishing moments.

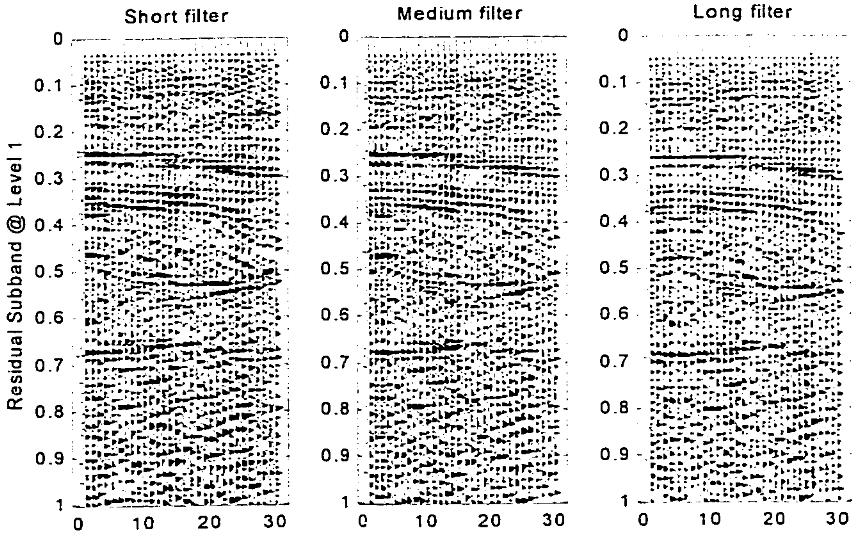

Short filter
Medium filter
Long filter

Figure 7-6: Symmetric extension versus the support width using the CDF(3.9) biorthogonal wavelet system.

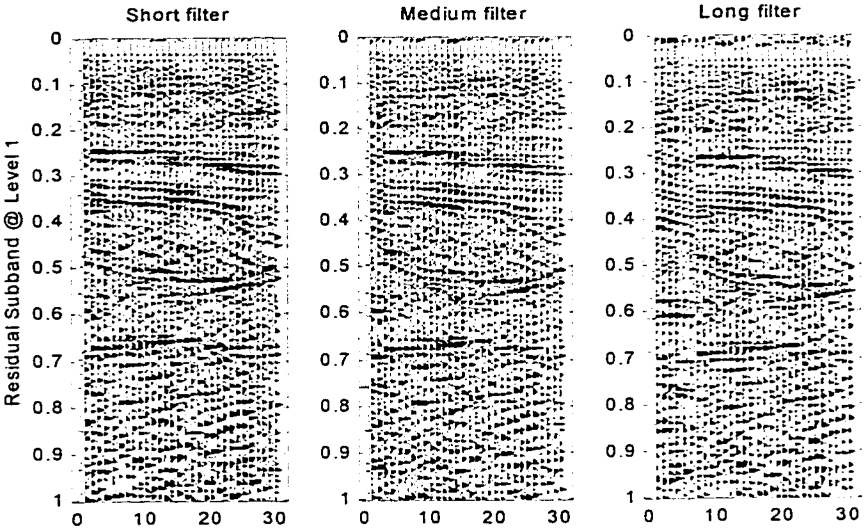


Figure 7-7: Periodic extension versus the support width using the CDF(3.9) biorthogonal wavelet system.

wavelet transform relies on the scaling and wavelet bases or their corresponding discrete filters. In this section, the effect of the orthogonality, the symmetry, and the smoothness of the bases is studied. This will allow to experimentally select the most adequate 2-D bases in order to perform a subsequent successful seismic compression.

Orthogonal bases

The compactly supported orthogonal wavelets have the maximum number of vanishing moment (VM), i.e. zeros at $z = -1$, for the minimum support width. In figure 7-8, the zeros of the daublet filters with respectively $p = 1, 2, 4, 8$ VM are illustrated, whereas figure 7-9 shows that the filters support $[2p - 1]$ is proportional the number of VM.

The most asymmetric 2-D daublet-2 wavelet system is shown in figure 7-10. The

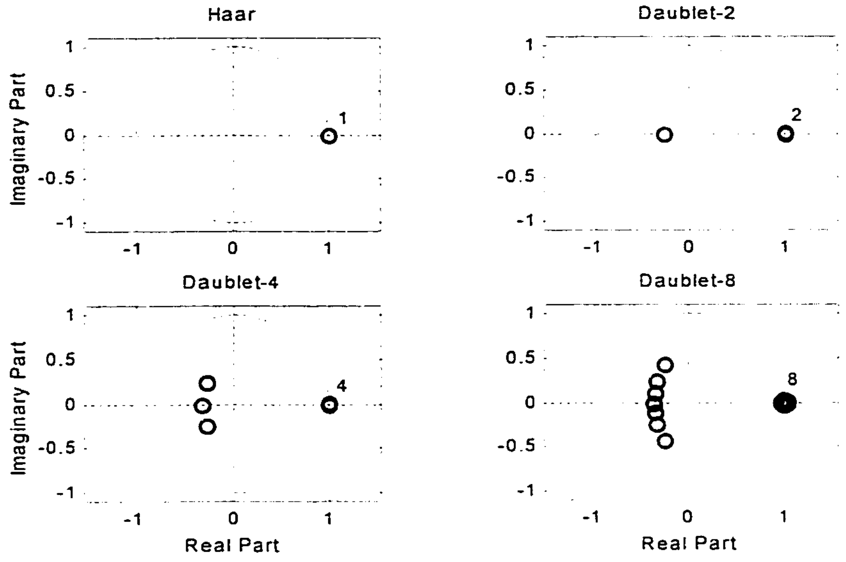


Figure 7-8: The number of vanishing moments p is proportional to the number of zeros at $z = -1$ of the wavelet filters.

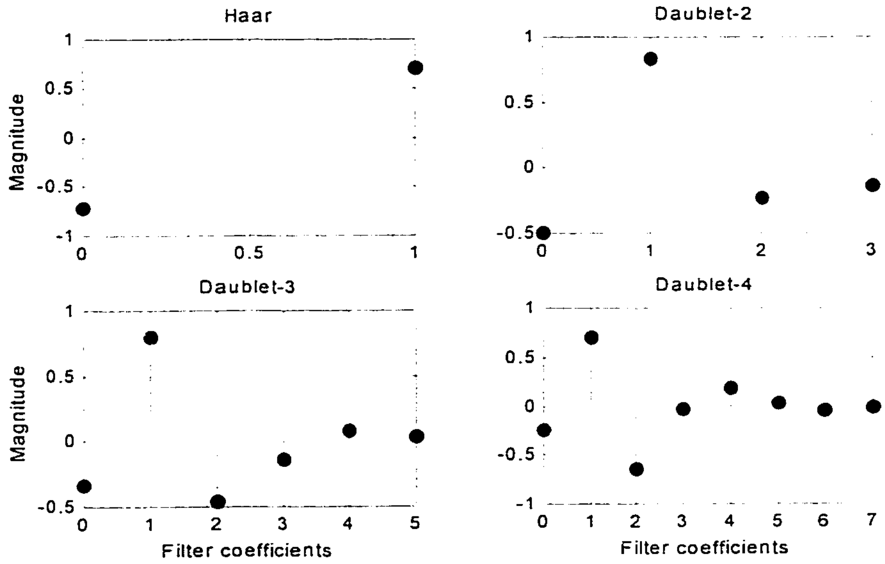


Figure 7-9: The wavelet filter support $[2p-1]$ is proportional to the number p of VM.

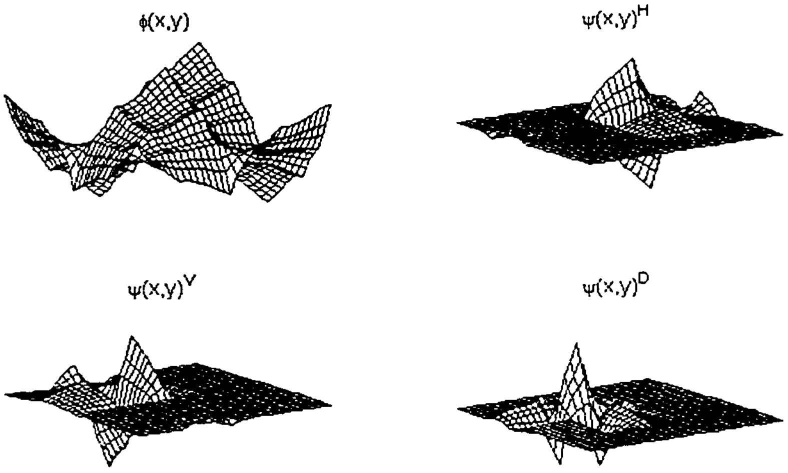


Figure 7-10: The separable 2-D daublet-2 anisotropic wavelets and scaling functions.

Fourier transforms of these bases, depicted in figure 7-11, confirm that the 2-D scaling function has a lowpass characteristic, whereas the three anisotropic 2-D wavelets reveal a highpass characteristic and a frequency selectivity along the vertical, horizontal and diagonal directions respectively. Furthermore, from the contour plots of figure 7-12, it could be inferred that, as stipulated by the *Heisenberg* principle, the decay of the bases in the frequency domain is slow because of their fast decay in the time domain.

Another member of the orthogonal wavelets with 8 VM, namely the daublet-8 system is depicted in 7-13. The bases have more smoothness and a much better frequency selectivity as illustrated by the *Fourier* transforms and their contour plots of figures 7-14 and 7-15 respectively. It can also be noticed that in addition to the overlap between the 2-D scaling and wavelet filters, the transition regions of the frequency responses of figure 7-12 are not steep. The filters overlap is much more pronounced than for the filters of figure 7-

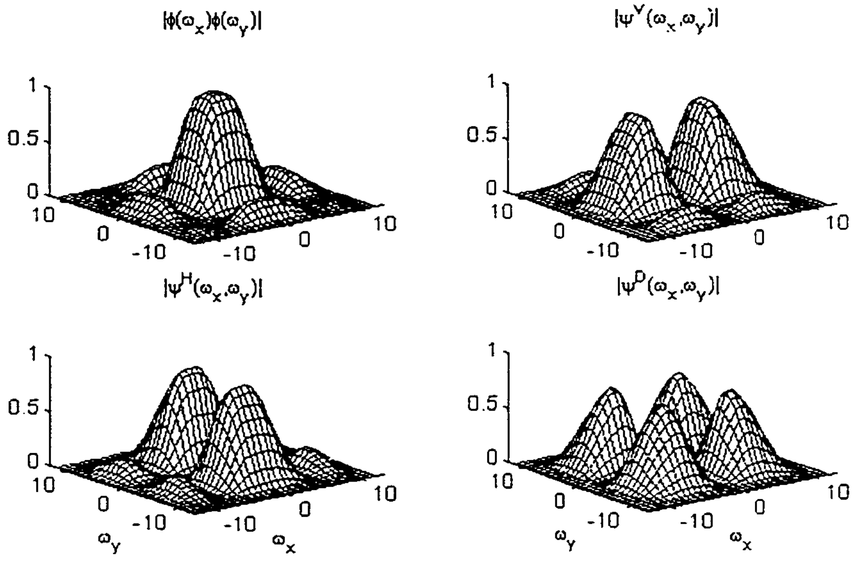


Figure 7-11: Modulus of the *Fourier* transform of the daublet-2 wavelet system.

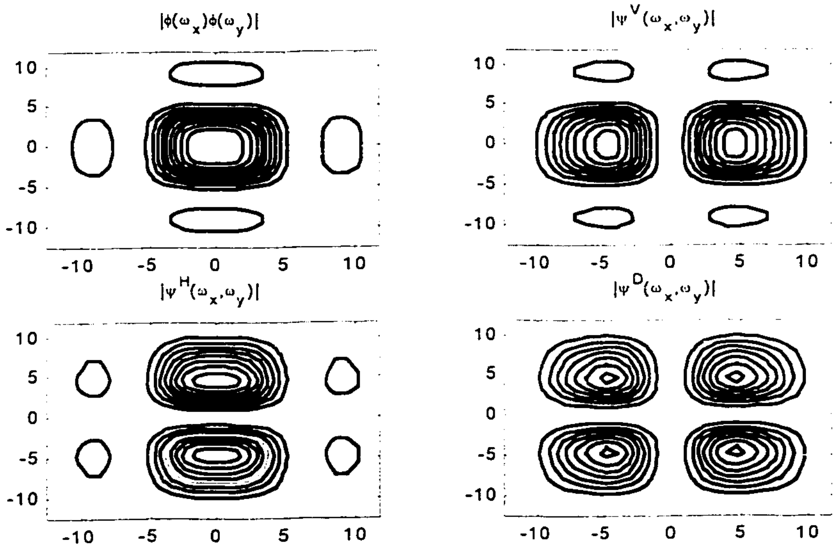


Figure 7-12: Contour plot of the modulus of the *Fourier* transform of the daublet-2 wavelet system.

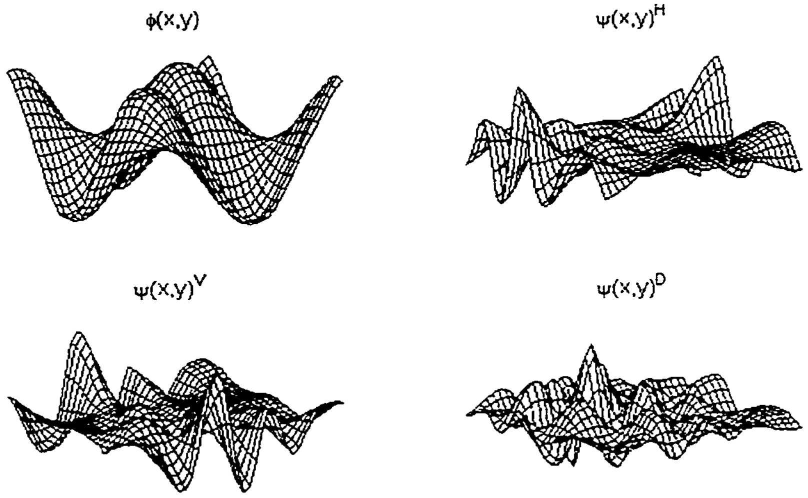


Figure 7-13: The separable 2-D daublet-8 anisotropic wavelets and scaling functions.

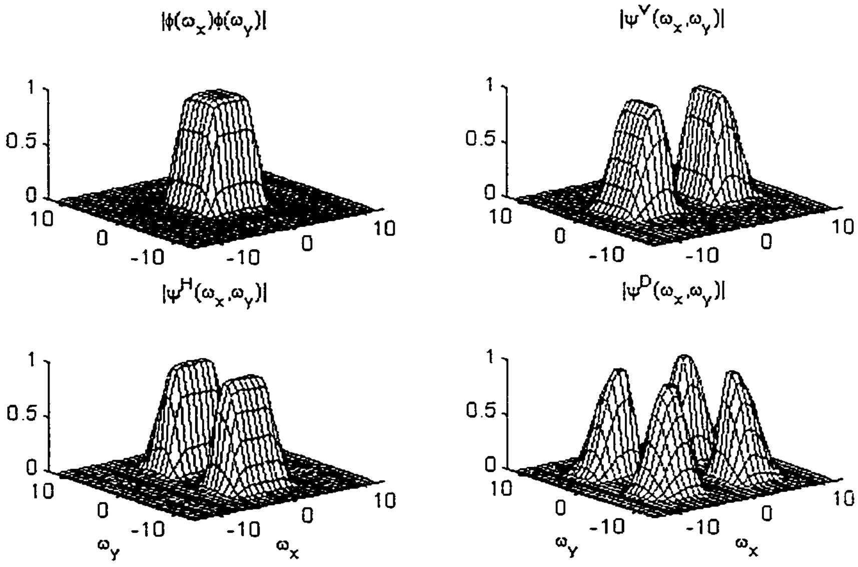


Figure 7-14: Modulus of the *Fourier* transform of the daublet-8 wavelet system.

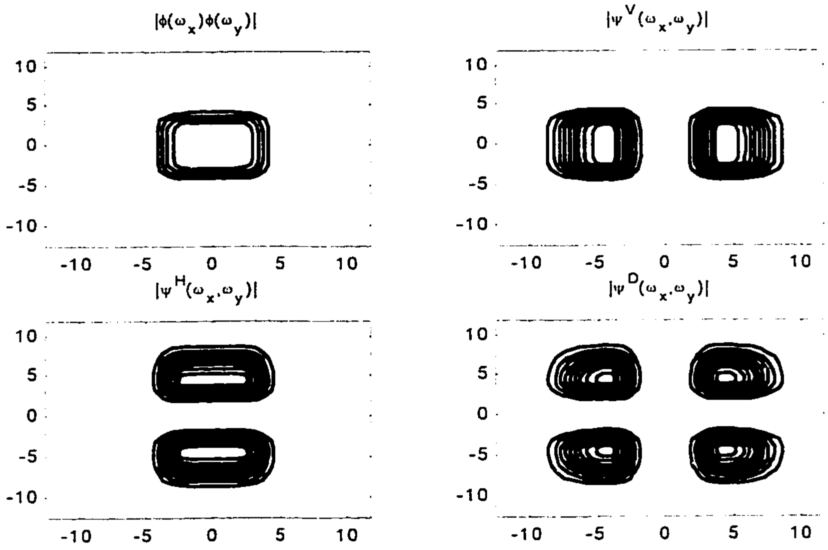


Figure 7-15: Contour plot of the modulus of the *Fourier* transform of the daublet-8 wavelet system.

15. Consequently, the crosscorrelation between the coefficients in the residual and the three detail subbands will not be zero. However, less crosscorrelation is expected with the daublet-8 filters. This fact is substantiated in figures 7-16 and 7-17 which compare the crosscorrelation functions (CCF) for some representative subbands obtained using the daublet-2 and daublet-8 bases respectively.

A three-level wavelet decomposition has been applied to the *Midyan* seismic image. This yields a residual subband whose size is 1/8 of the original size of the data and 9 detail subbands at different scales and orientations with various sizes, i.e. $1/2, 1/4$ and $1/8$ of the original size. The residual subband bears the gross structure of the original data as it is only subject to lowpass filtering. The details subbands contain as their part, the extracted details along the three orientations, i.e., vertical, horizontal and diagonal. The complete

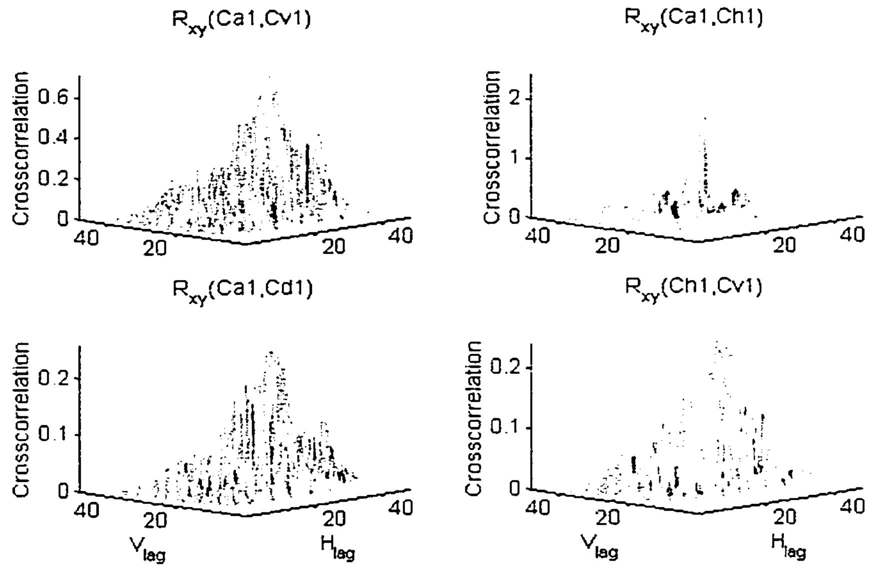


Figure 7-16: Crosscorrelation functions for some representative subbands at the first level of the wavelet decomposition using the daublet-2 system.

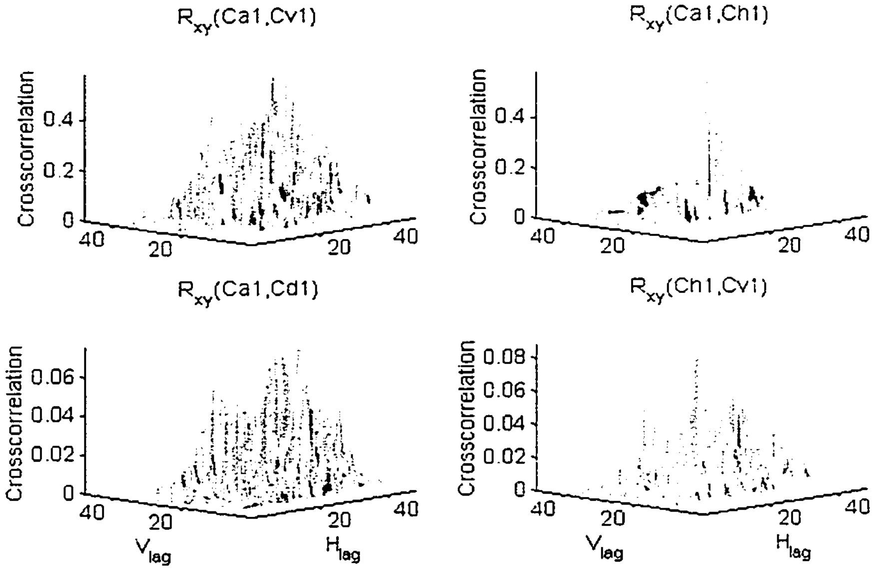


Figure 7-17: Crosscorrelation functions for some representative subbands at the first level of the wavelet decomposition using the daublet-8 system.

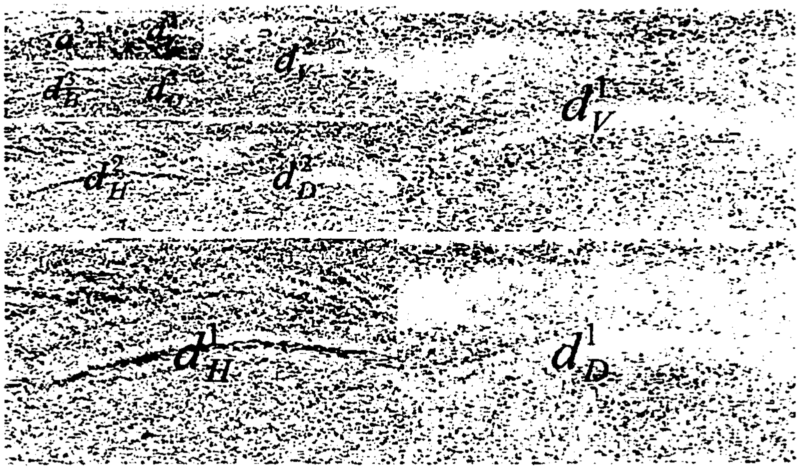


Figure 7-18: 3-level daublet-2 wavelet expansion coefficients arranged in *Mallat* format.

wavelet decomposition is usually displayed in the *Mallat* format as illustrated with figure 7-18 for the daublet-2 wavelet system. The essence of decomposing the seismic data using anisotropic 2-D wavelet bases is to extract the relevant structure along the three directions such that only few coefficients will be significant while the rest will be negligibly small. The VDD mode of figure 7-18, although not providing a clear view because of the decreasing density of the wavelet coefficients in the subbands, shows how the details are extracted especially along the horizontal details. When compared to the wavelet decomposition of figure 7-19, which uses a wavelet system with more vanishing moments, the magnitude of the coefficients is smaller and even there are a lot of them close to zero. This confirms again that the daublet-8 system yields a better decorrelation of the original seismic data.

The wavelet transform is also expected to reduce the entropy of the original data. In figure 7-20, it is demonstrated how the entropy is reduced using the normalized histograms

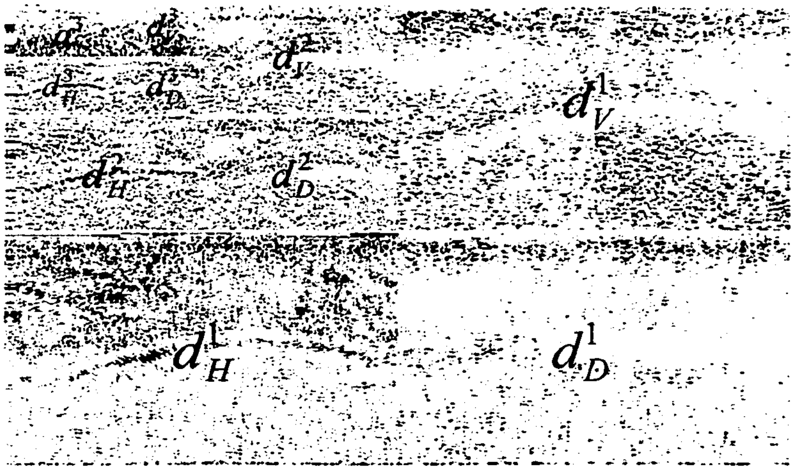


Figure 7-19: 3-level daublet-8 wavelet expansion coefficients arranged in *Mallat* format.

with respect to the original data. When compared to the histograms of figure 7-21 that are obtained using the daublet-8 bases, it ensues that the entropy is much more reduced at the first scale especially in the horizontal and diagonal directions. This is an indicator that most of the information present in the original data lie in both the horizontal and diagonal directions, which represent respectively the relatively flat reflectors of the clastics and the bending parts of the salt diapirism and faults. These assumptions are confirmed by inspecting the reconstructed subband at the first scale and along the three orientations using the VAD plots of figures 7-22, 7-23, 7-24 and 7-25. These plots also compare the performance of the two wavelet systems in terms of smoothness of the bases and their frequency selectivity. Indeed the reconstructed traces using the daublet-2 system are more spiky because of the high irregulatity of the bases. Moreover, the slow decay of the latter in the frequency domain is responsible of the failure of the transform to properly

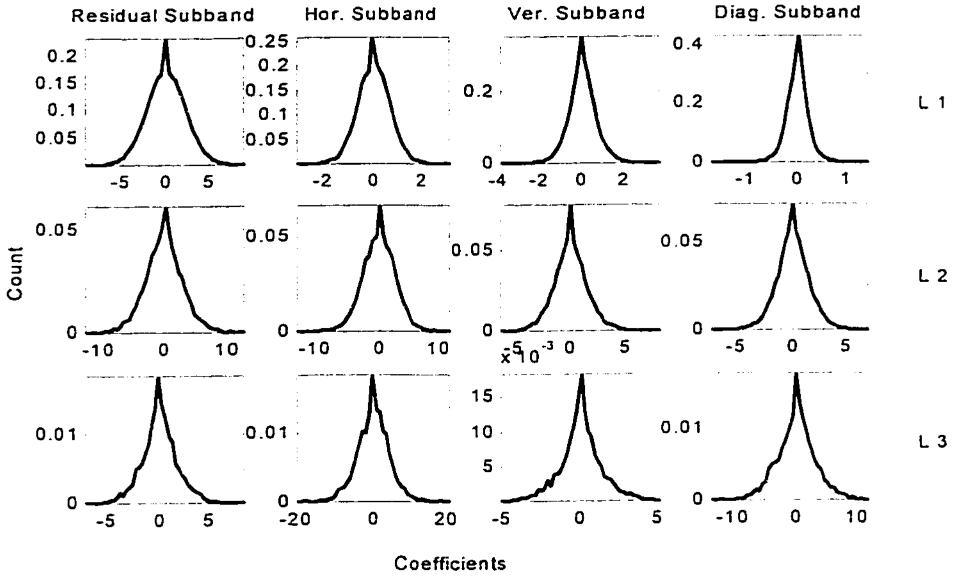


Figure 7-20: Normalized histograms of the wavelet coefficients at all the scales and orientations using the daublet-2 system.

extract the horizontal and diagonal details respectively. This is also noticeable from the degree of correlation estimated using the ACF and their contour plots of figures 7-26 and 7-27 for the coefficients lying in the horizontal and diagonal subbands.

Biorthogonal bases

The B-splines biorthogonal wavelet bases constitute another important class of modern wavelet systems. The biorthogonal wavelets are symmetric and use different FIR filters pairs for the decomposition and reconstruction stages. The performance of two biorthogonal system is investigated. These are the CDF(3,9) wavelet system, which has $N = 3$ real moments and $\tilde{N} = 9$ dual moments, and the CDF(4,4) wavelet system with $N = \tilde{N} = 4$, also known as the FBI wavelet system. This choice is dictated by the ability of each

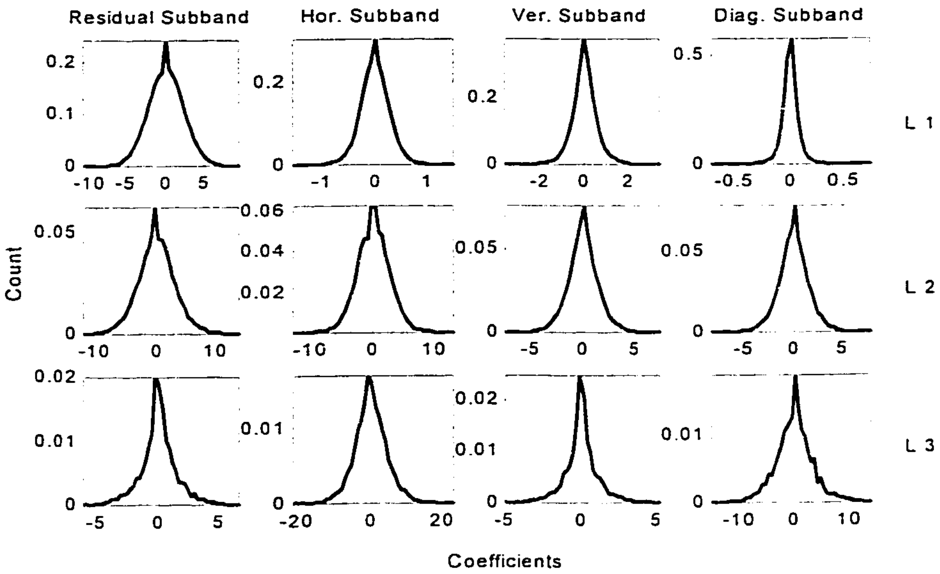


Figure 7-21: Normalized histograms of the wavelet coefficients at all the scales and orientations using the daublet-8 system.

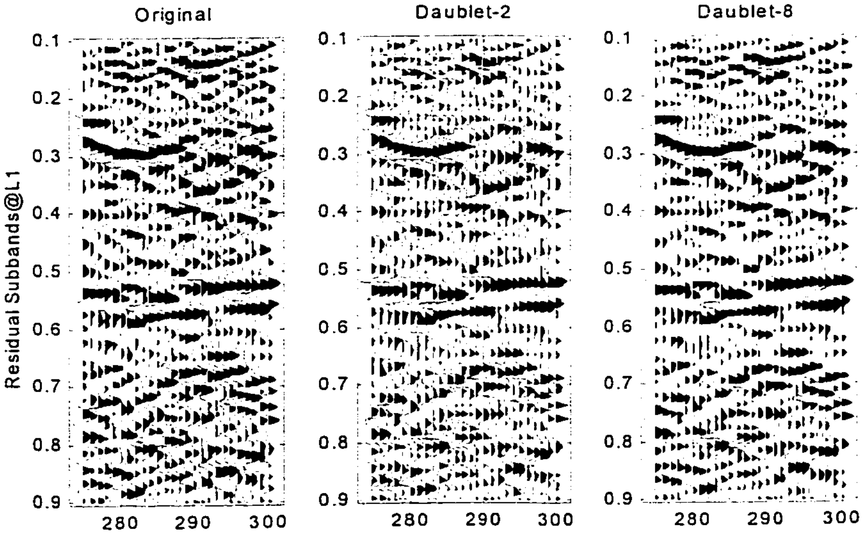


Figure 7-22: Original data versus the first level residual subbands using the daublet-2 and daublet-8 wavelet systems.

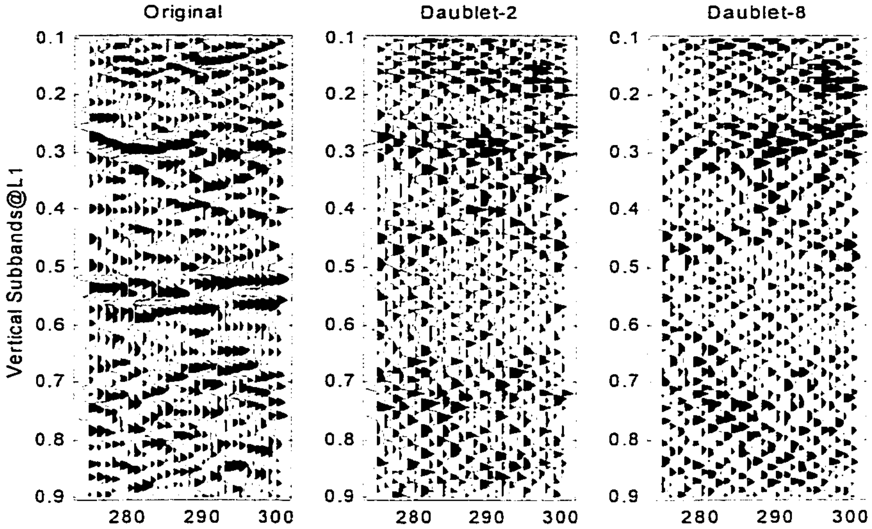


Figure 7-23: Original data versus the first level vertical subbands using the daublet-2 and daublet-8 wavelet systems.

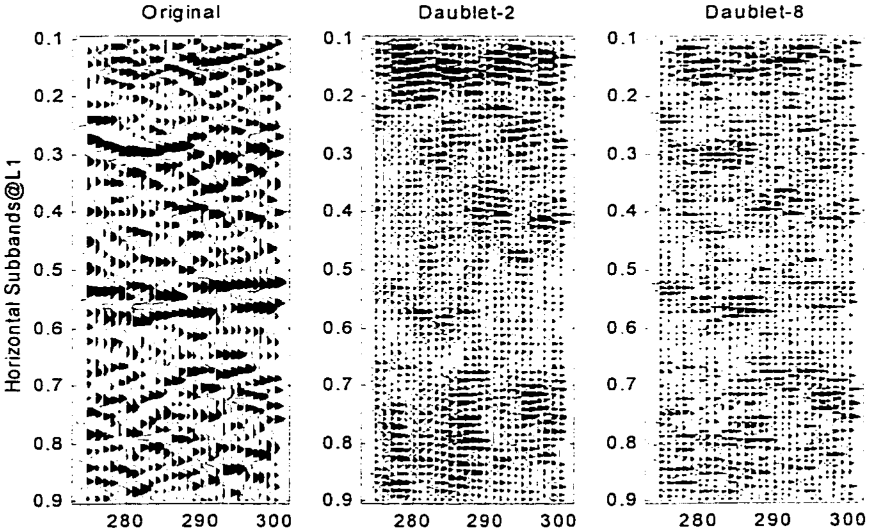


Figure 7-24: Original data versus the first level horizontal subbands using the daublet-2 and daublet-8 wavelet systems.

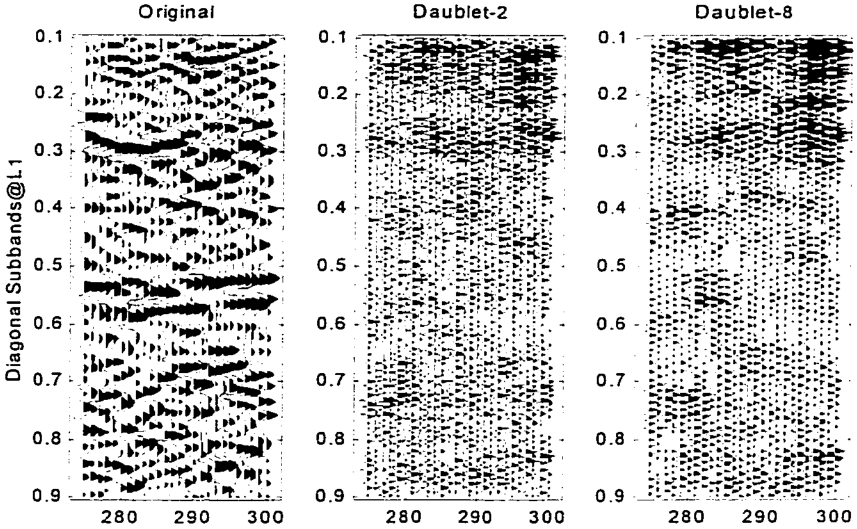


Figure 7-25: Original data versus the first level diagonal subbands using the daublet-2 and daublet-8 wavelet systems.

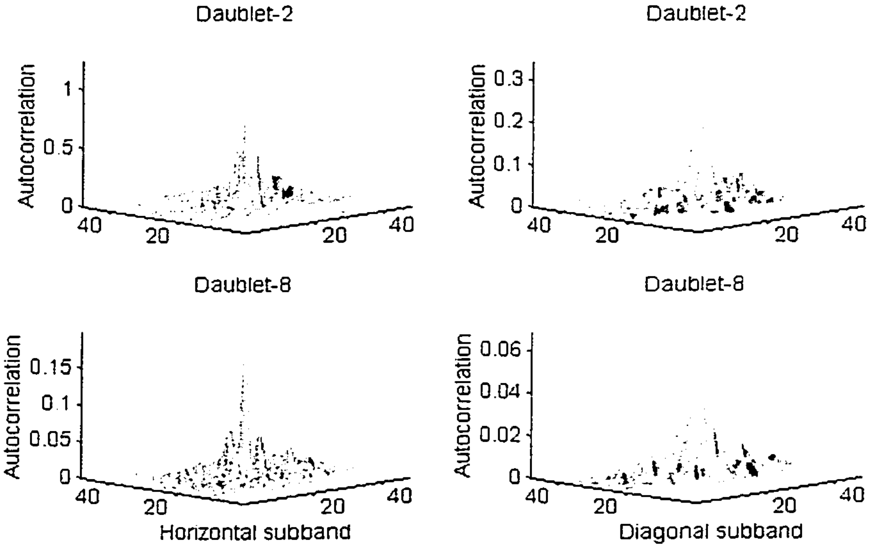


Figure 7-26: Autocorrelation functions of the first level horizontal and diagonal subbands using daublet-2 and daublet-8 systems.

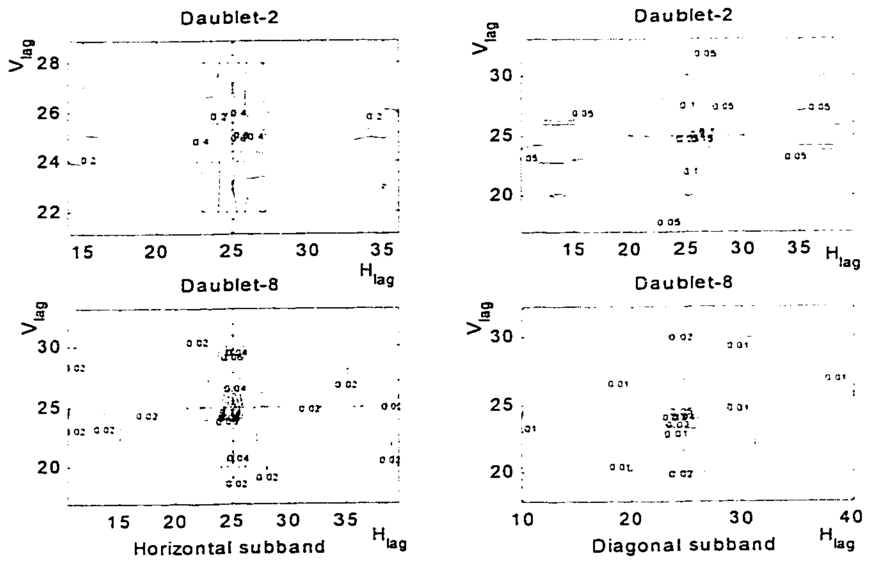


Figure 7-27: Contour plots of the ACF of the first level horizontal and diagonal subbands using daublet-2 and daublet-8 systems.

wavelet system to compact the energy as demonstrated in figure 7-28. It can be clearly seen that the CDF(3,9) system confines 65% of the total data energy to 5000 wavelet coefficients against ~50% for the CDF(4,4) and the daublet-45 systems. It is also worthy to notice that we only get a marginal gain in energy, i.e. ~5% when increasing the number of VM from 2 to 45 for the daublets. This confirms the trade-off between the number of VM and the size of the wavelet filters. In figure 7-29, the characteristics of the 1-D filters of the CDF(3,9) wavelet system are represented. In order to speed up the wavelet expansion, the shortest wavelet filter is used at the analysis stage. On the contrary, the most regular one is used at the synthesis stage in order to smooth the compression errors. From the 2-D analysis bases and their spectra displayed respectively in figures 7-30 and 7-31, it can be inferred that the bases are smoother than the daublet bases for the same

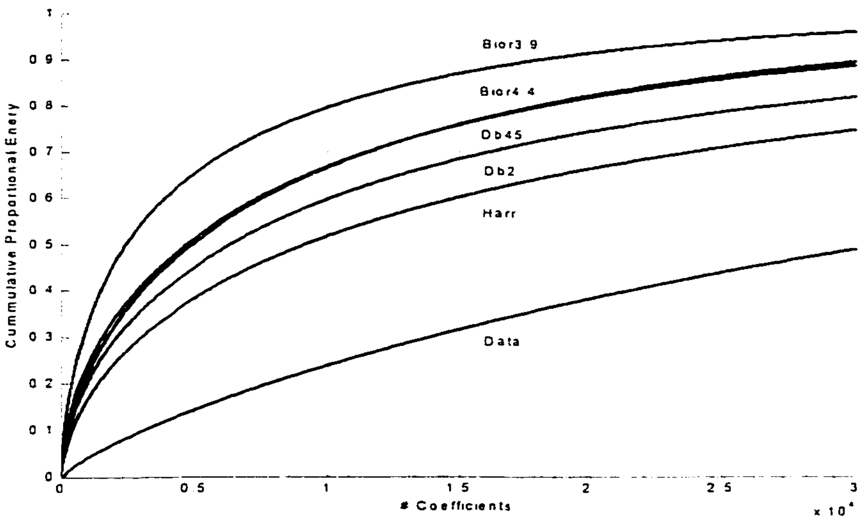


Figure 7-28: Performance of orthogonal and biorthogonal bases in compacting the energy of the input data.

number of VM. The frequency spectrum confirms also that the filters are symmetric and have a fast frequency decay. For the sake of comparison, the 2-D bases and their spectra are displayed in figures 7-32 and 7-33 respectively. Again, the bases are much smoother than those of the daublet system. This gives a forerunner evidence of the superiority of the biorthogonal wavelet system over the orthogonal one. The result of the three-level biorthogonal wavelet decomposition appears in figure 7-34.

Next, the performance of the CDF(3,9) system is compared to the FBI one. For the FBI system, the scaling and wavelet filter have approximately the same length, i.e. 7 and 9 taps. The decomposition and reconstruction wavelets have the same number of vanishing moments $N = \tilde{N} = 4$ as shown in figure 7-35. Again, in order to ensure a fast decomposition and a smooth reconstruction of the data, the shorter lowpass filter

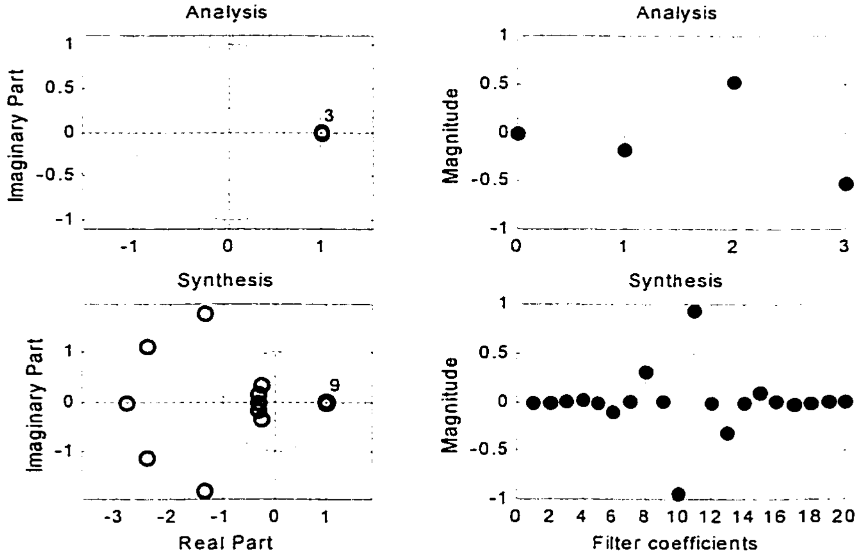


Figure 7-29: Zeros and coefficients for the CDF(3,9) wavelet system filters.

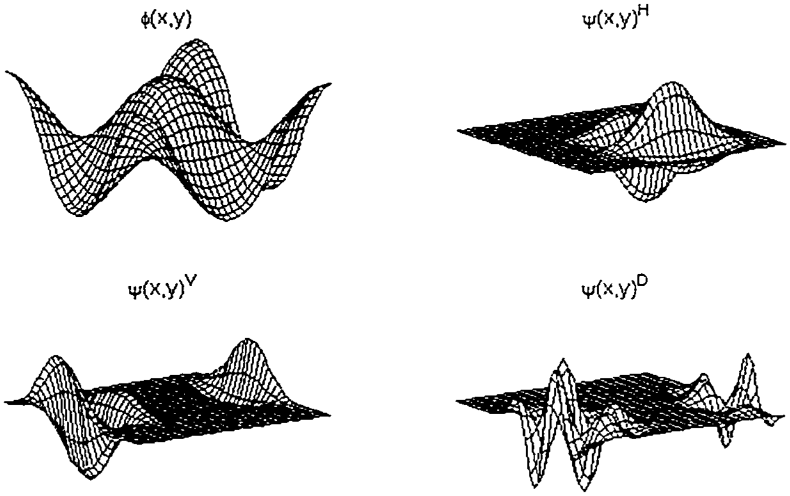


Figure 7-30: The separable 2-D analysis CDF(3,9) wavelets and scaling functions.

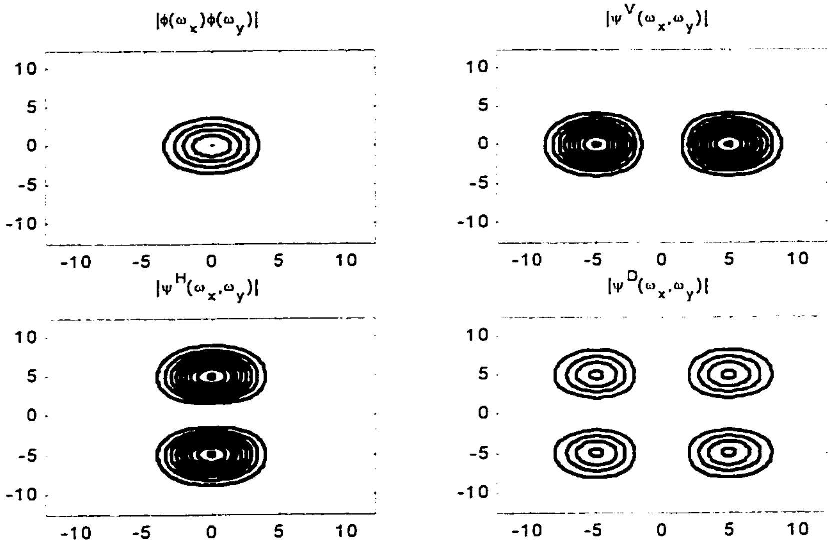


Figure 7-31: Contour plot of the modulus of the *Fourier* transform of the 2-D analysis CDF(3,9) bases.

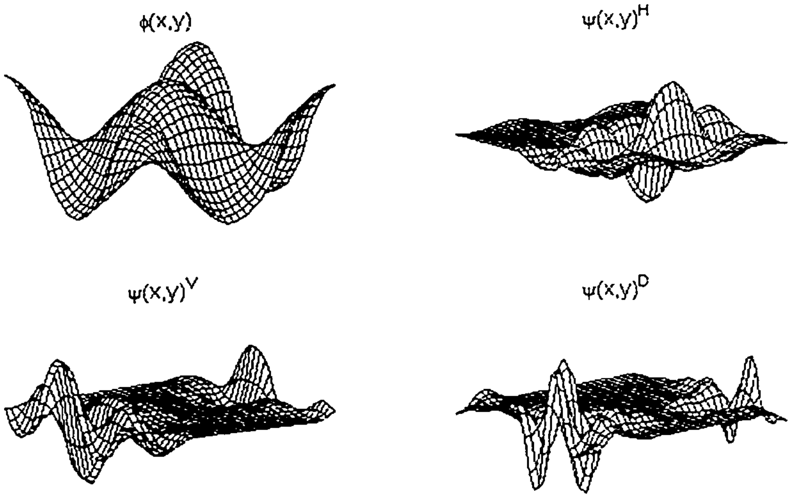


Figure 7-32: The separable 2-D synthesis CDF(3,9) wavelets and scaling functions.

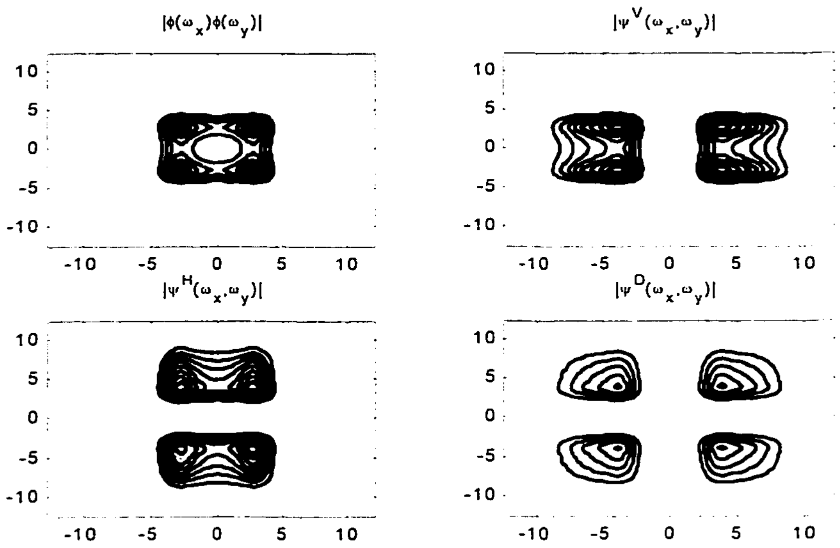


Figure 7-33: Contour plot of the modulus of the *Fourier* transform of the 2-D synthesis CDF(3,9) bases.

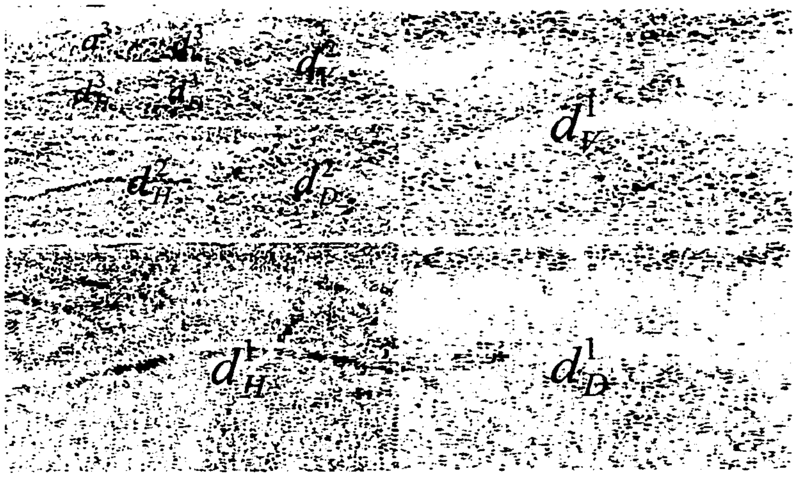


Figure 7-34: 3-level CDF(3,9) wavelet expansion coefficients arranged in *Mallat* format.

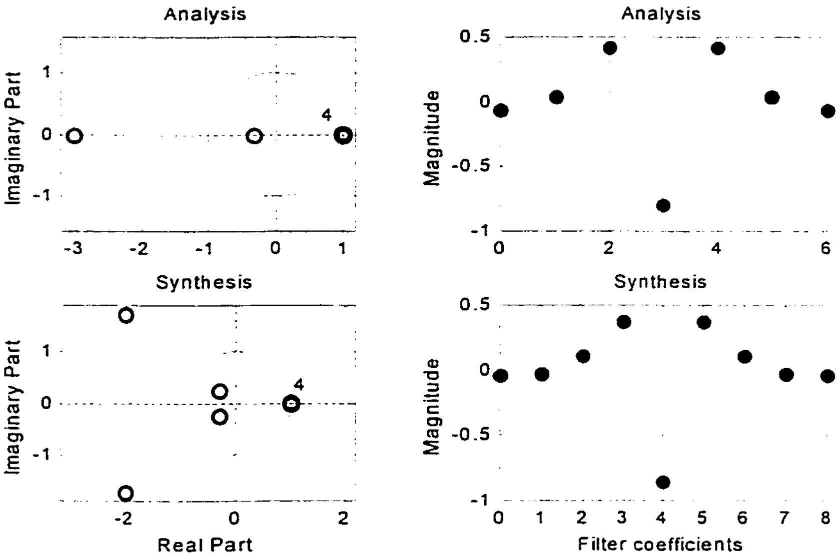


Figure 7-35: Zeros and coefficients for the analysis CDF(4,4) wavelet system filters.

is used at the decomposition stage whereas the longer one is used at the reconstruction stage. The 2-D analysis bases and their *Fourier* transforms are displayed in figures 7-36 and 7-37 respectively. The result of the three-level wavelet decomposition arranged in *Mallat* format is displayed in figure 7-38 respectively. To get a better evaluation of the performance of the two biorthogonal wavelet systems, a thorough comparative study is performed using the reconstructed subbands at all the scales and orientations. Using a VAD mode that allows a better inspection of the waveforms and the reflectors, the results are plotted in figures 7-39, 7-40, 7-41 and 7-25 at the first scale, in figures 7-43, 7-44, 7-45 and 7-45 at the second scale and in figures 7-46, 7-47, 7-48 and 7-49 at the coarsest scale.

The visual inspection of the waveforms reveals that the CDF(3,9) performs better than the CDF(4,4) wavelet system. The waveforms are in general more spiky in the case of the CFD(4,4) system. Moreover, the two reflectors between 0.5 and 0.6 sec are well preserved

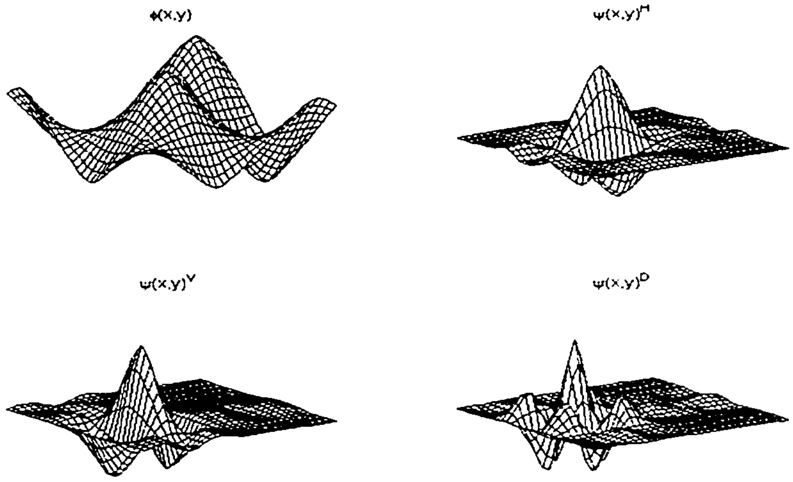


Figure 7-36: The separable 2-D analysis CDF(4,4) wavelets and scaling functions.

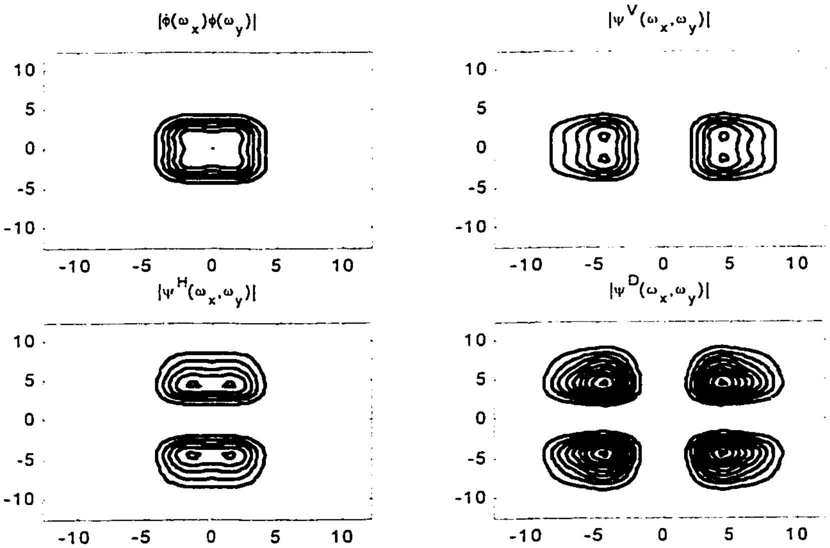


Figure 7-37: Contour plot of the modulus of the *Fourier* transform of the 2-D analysis CDF(4,4) bases.

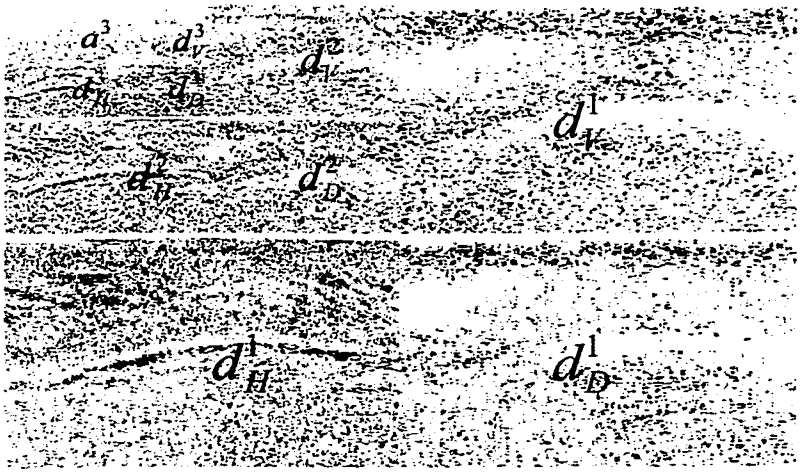


Figure 7-38: 3-level biorthogonal (4,4) wavelet expansion coefficients arranged in *Mallat* format.

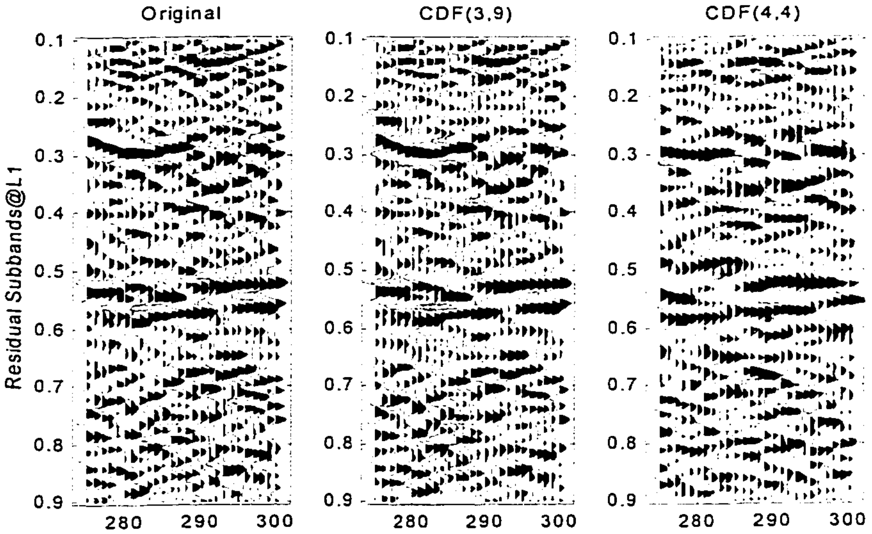


Figure 7-39: Original data versus the first level residual subbands using the CDF(3,9) and CDF(4,4) wavelet systems.

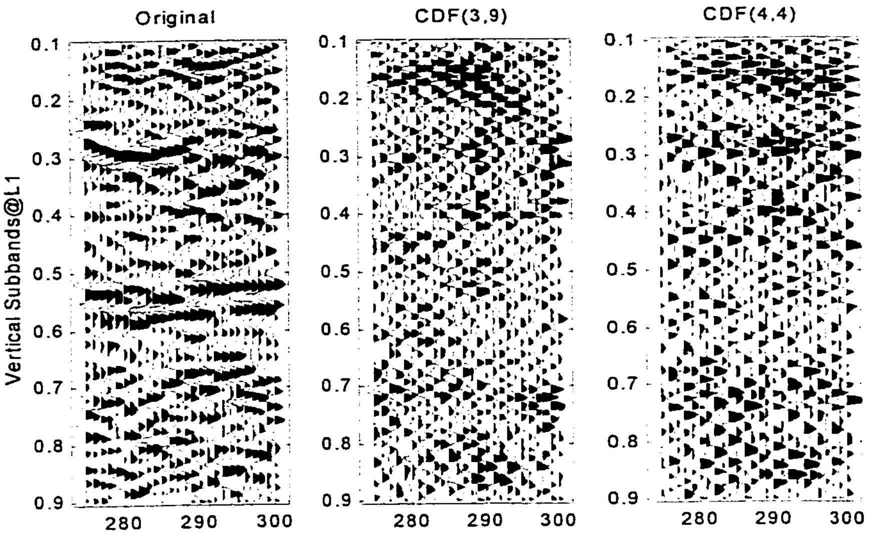


Figure 7-40: Original data versus the first level vertical subbands using the CDF(3,9) and CDF(4,4) wavelet systems.

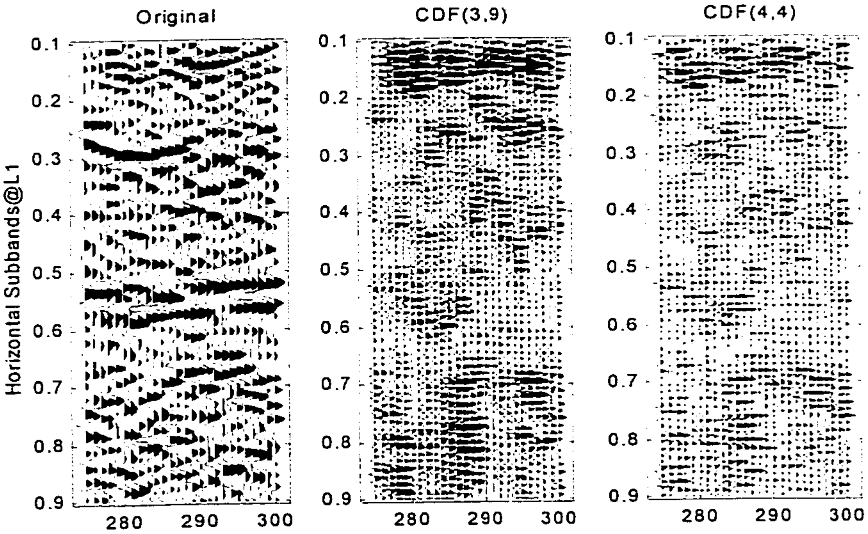


Figure 7-41: Original data versus the first level horizontal subbands using the CDF(3,9) and CDF(4,4) wavelet systems.

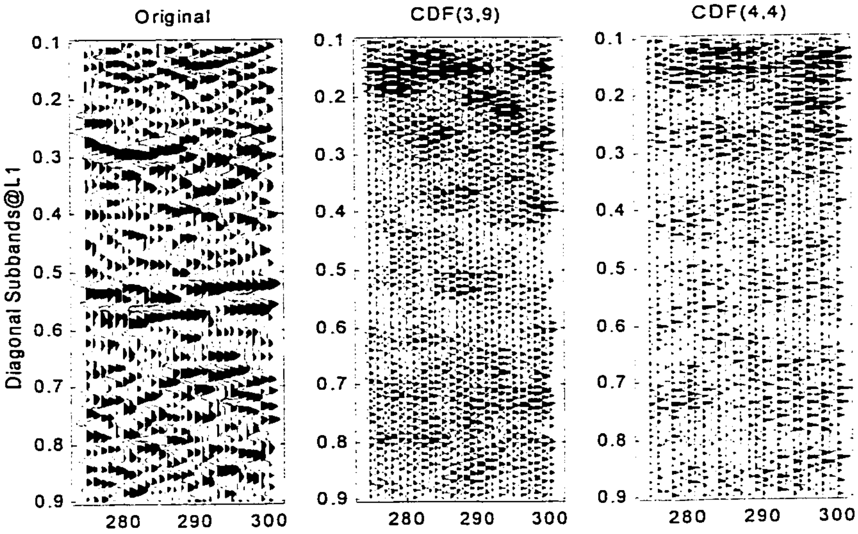


Figure 7-42: Original data versus the first level diagonal subbands using the CDF(3,9) and CDF(4,4) wavelet system.

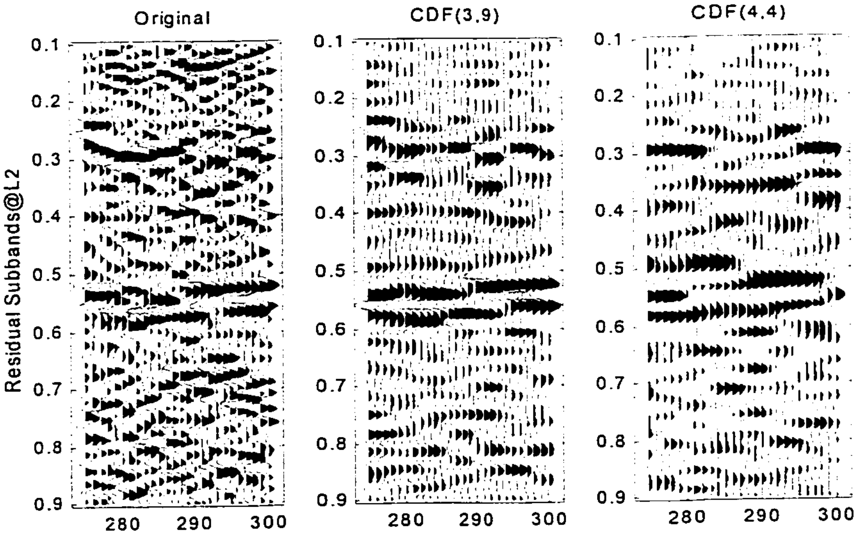


Figure 7-43: Original data versus the second level residual subbands using the CDF(3,9) and CDF(4,4) wavelet systems.

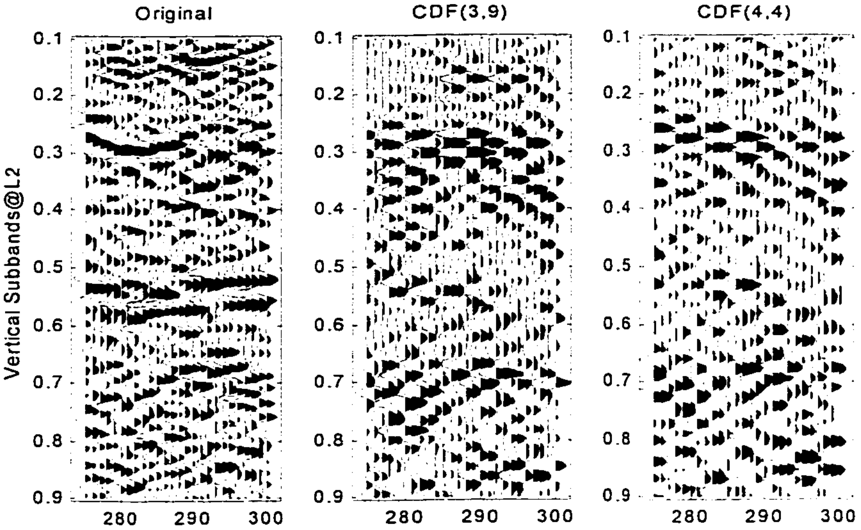


Figure 7-44: Original data versus the second level vertical subbands using the CDF(3,9) and CDF(4,4) wavelet systems.

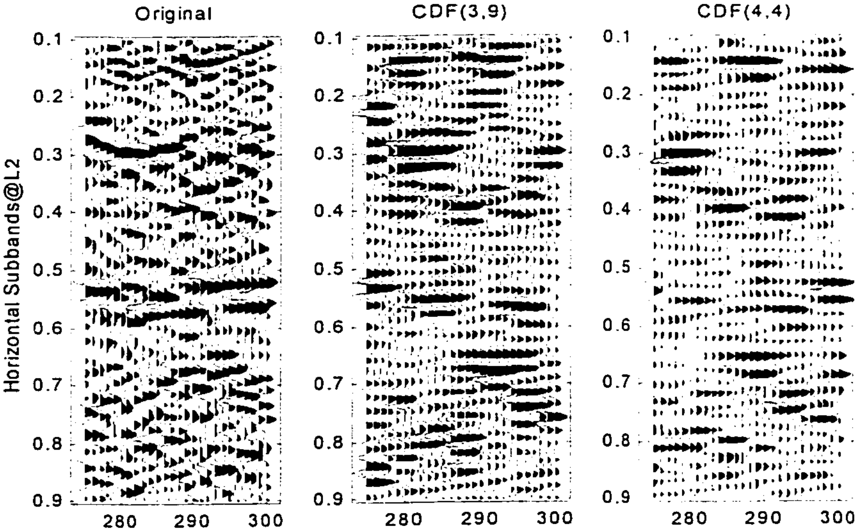


Figure 7-45: Original data versus the second level horizontal subbands using the CDF(3,9) and CDF(4,4) wavelet systems.

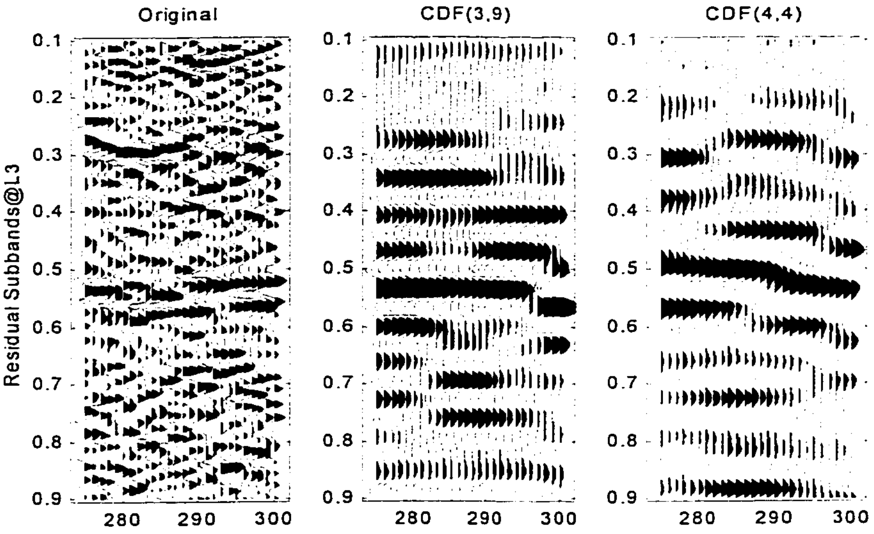


Figure 7-46: Original data versus the third level residual subbands using the CDF(3,9) and CDF(4,4) wavelet systems.

in the residual subbands with the CDF(3,9) system as it clearly appears in figure 7-39. Elsewhere, the continuous reflector starting above 0.3 sec appears as discontinuous in the case of the CDF(4,4), which may generate visual artifacts in the reconstructed data. Figure 7-46 clearly shows that most of the reflectors are well reproduced in the coarsest residual subband with the CDF(3,9) system. This is not the case with the CDF(4,4), where even some flat reflectors look like dipping ones. Notice that in figure 7-48, the weak events around 0.2 and 0.8 sec are almost inexistent.

To end this section, which aims at selecting the best wavelet bases for subsequent compression, the intra-band correlations are plotted in figures 7-50, 7-52, 7-52 and 7-53. The interband correlation for some representative subbands is also plotted in figure 7-54. The results are self-explanatory and confirm that the CDF(3,9) wavelet system performs

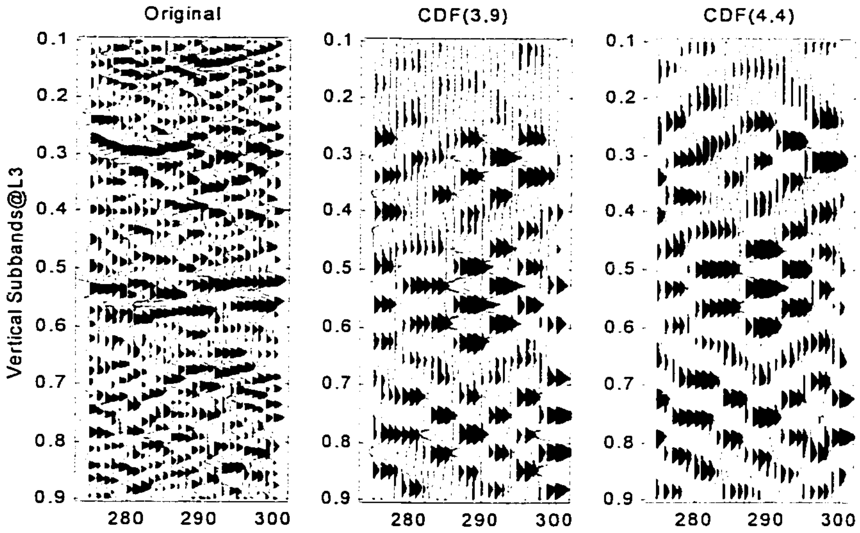


Figure 7-47: Original data versus the third level vertical subbands using the CDF(3,9) and CDF(4,4) wavelet systems.

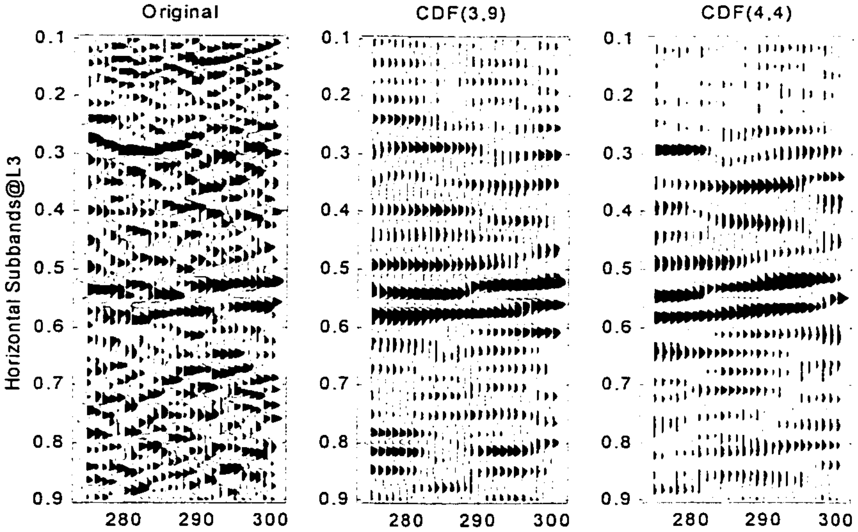


Figure 7-48: Original data versus the third level horizontal subbands using the CDF(3,9) and CDF(4,4) wavelet systems.

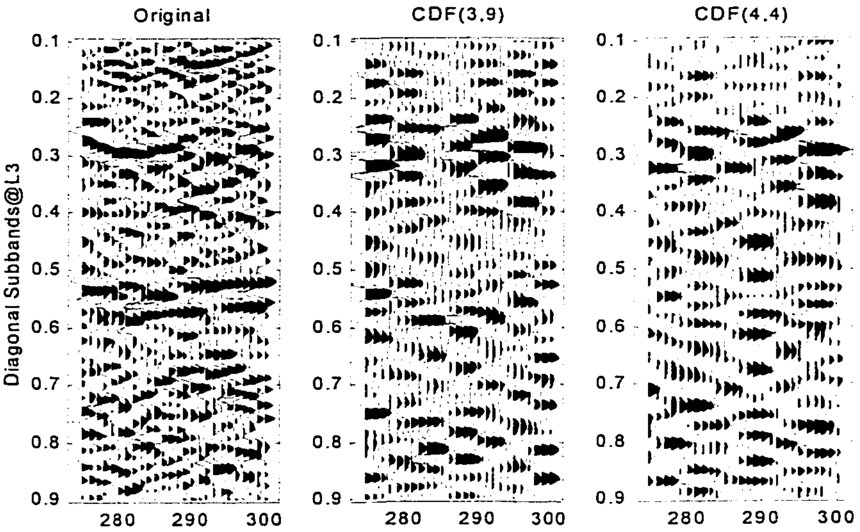


Figure 7-49: Original data versus the third level diagonal subbands using the CDF(3,9) and CDF(4,4) wavelet systems.

better than the FBI one.

7.3 Wavelet Compression

7.3.1 Metrics

Data quality metrics In comparing an original 2-D signal $f(x,y)$ to an inexact reconstruction $\check{f}(x,y)$, the most common error measure in the engineering field is the mean-squared error (MSE):

$$MSE = \frac{1}{MN}\left\|f-\check{f}\right\|_2^2 = \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\left|f(i,j)-\check{f}(i,j)\right|^2 \tag{7.1}$$

In image processing applications such as compression, a logarithmic metric based on

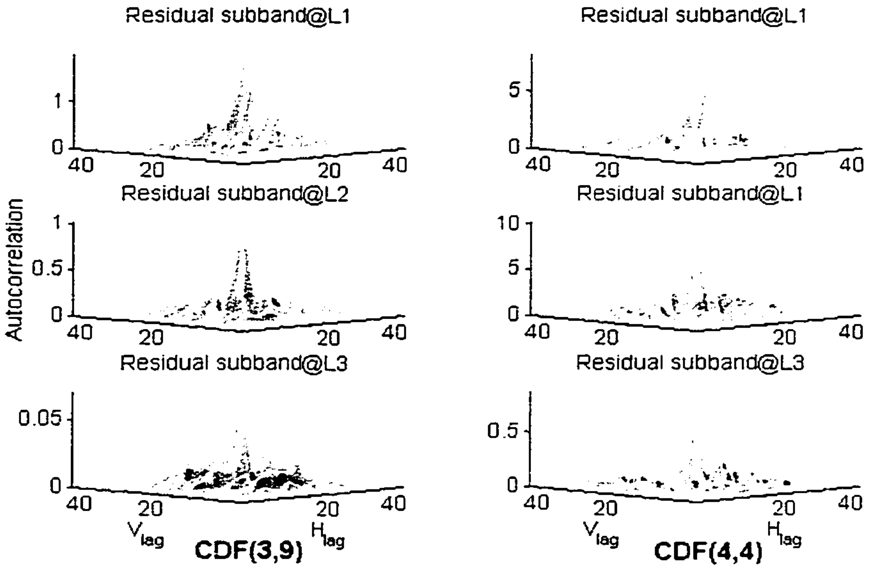


Figure 7-50: Comparison of the coefficients correlation in the residual subbands for the CDF(3.9) and CDF(4,4) wavelet systems.

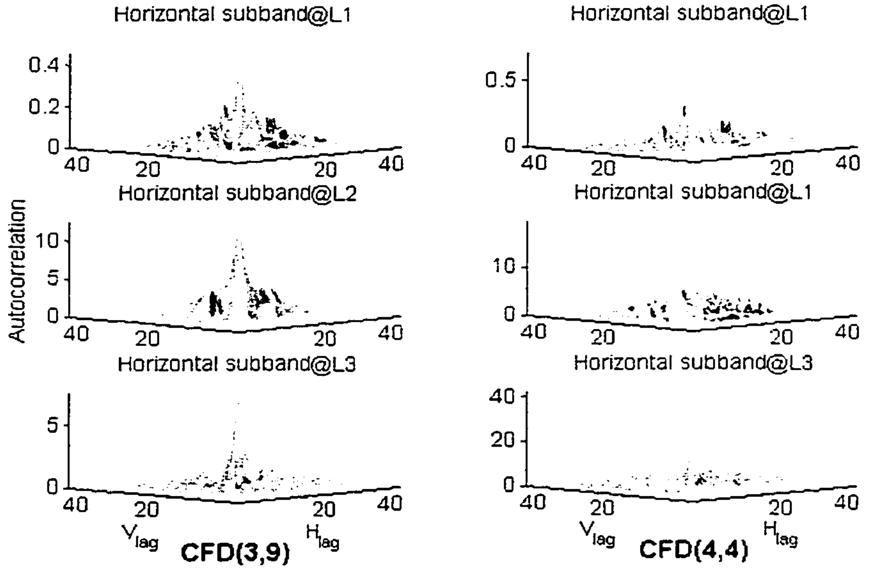


Figure 7-51: Comparison of the coefficients correlation in the horizontal subbands for the CDF(3,9) and CDF(4,4) wavelet systems.

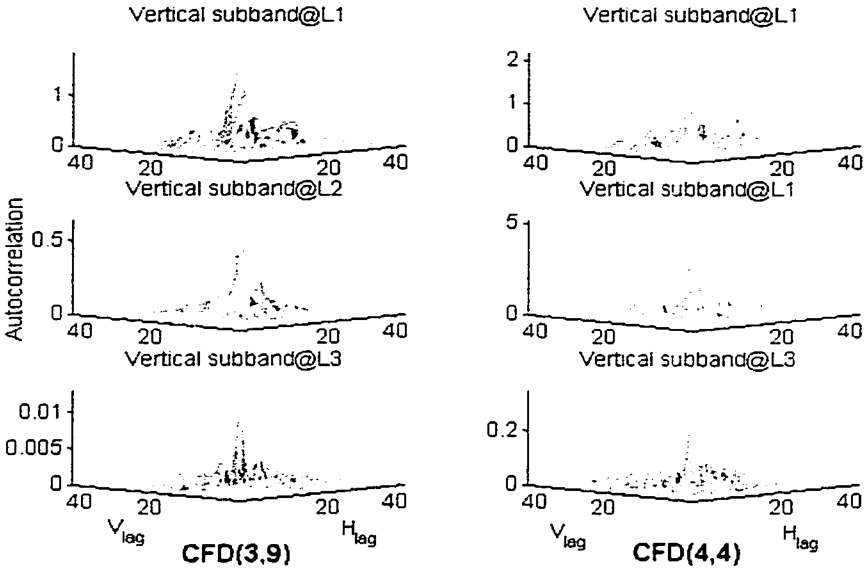


Figure 7-52: Comparison of the coefficients correlation in the horizontal subbands for the CDF(3,9) and CDF(4,4) wavelet systems.

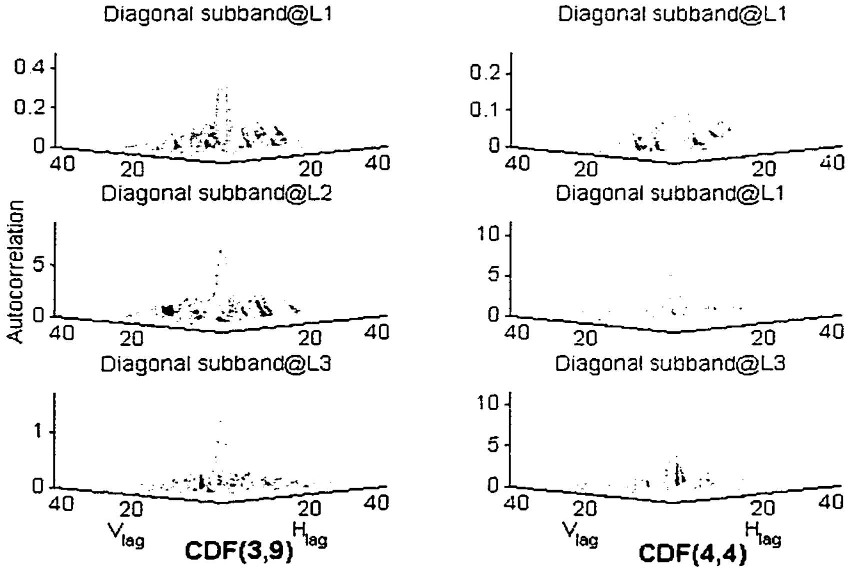


Figure 7-53: Comparison of the coefficients correlation in the diagonal subbands for the CDF(3,9) and CDF(4,4) wavelet systems.

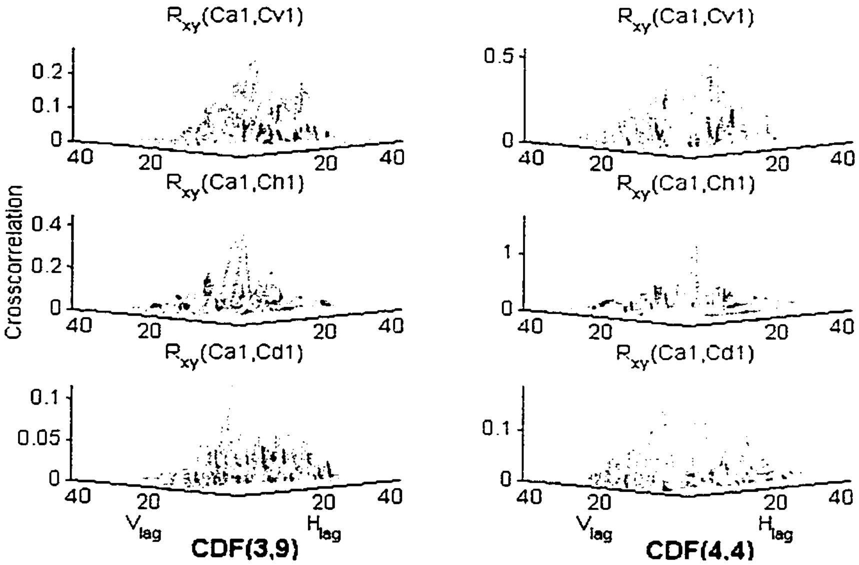


Figure 7-54: Comparison of the inter-band correlation of some representative subbands for the CDF(3,9) and CDF(4,4) wavelet systems.

the MSE is often used. It is called peak signal-to-noise ratio ($PSNR$) and is given by

$$PSNR = 10\log_{10}\frac{\max|f(i,j)|}{MSE} \tag{7.2}$$

This error metric is commonly used because it is easy to compute and it is mathematically tractable. Elsewhere, it has a reasonable though imperfect correspondence[2] with image degradations as interpreted subjectively by human observers. This is also true for seismic data.

Due to thresholding and quantization, the compression/decompression process intro-

---

[2] At medium and low rates, the PSNR does not agree with the quality of the image.

duces an error, which can be expressed as

$$\epsilon_{ij} = f_{ij} - \check{f}_{ij}, \quad i = 1, ..., N \;\; ; \;\; j = 1, ..., M \tag{7.3}$$

where the size of the original data is NxM. Two aspects of this error are of interest, namely the total error and the correlation of the error. The quantification of the error is realized using the L^p norm, defined in the discrete case as

$$L^p = \left(\frac{\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \left| f(i,j) - \check{f}(i,j) \right|^p}{\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} |f(i,j)|^p} \right)^{1/p} x100 \tag{7.4}$$

where the error is normalized with the norm of the original data. The most used norms are L^1, L^2, and L^∞, which respectively estimate the total absolute error, the energy of the error and the maximum error.

The characterization of the compression error is obtained by inspecting the autocorrelation function (ACF). The ideal situation occurs when the ACF is a δ-distribution, which is an indictor of the complete decorrelation of the error samples. In the practical case, the ACF is not exactly a δ-distribution. Therefore, rather than trying to characterize the latter, it is sufficient to analyze the magnitude of the peak and the extent of the ACF.

In addition, VDD and VAD mode plots are used to visually evaluate the quality of the reconstructed data and the difference sections.

Compression scores

In the proposed strategies, compression occurs after the thresholding and after the quantization and encoding. Therefore, the adopted compression scores rely on both the retained energy after thresholding. i.e.,

$$E = 100\frac{\|\text{kept coeff. }\|}{\|\text{Total coeff.}\|} \tag{7.5}$$

and the compression ratio after encoding

$$CR = \frac{\text{Size of the compressed data (Mb)}}{\text{Original size (Mb)}} \tag{7.6}$$

7.3.2 Compression results

Coefficients distributions The statistical properties of the survived detail coefficients play a key role in the design of the TUSQ outside the threshold zone. This issue is treated experimentally using both the Q-Q plots and the histogram fitting technique, where the histogram of each detail subband coefficients is compared to the *Lapacien* pdf. The Q-Q plot is an effective statistic tool for verifying that the detail coefficients are close to the assumed distribution. Let $x_1, x_2, ..., x_N$ be the sample value of $x \sim Lap(\alpha)$. They are first ordered by increasing magnitude denoted by $y_1, y_2, ..., y_N$, called the order statistics of the absolute values. When $\alpha = 1; y = |x| \sim e^{-y}$, the cumulative distribution function is $F(a) = \int_0^a e^{-y}dy = q$ and $a = F^{-1}(q)$ is called the $q - th$ quantile, whence $a = -\ln(1-q)$. The Q-Q plot graphs the pairs $(y_k, -\ln(1-\frac{k}{N}), k = 1, 2, ..., N$. The

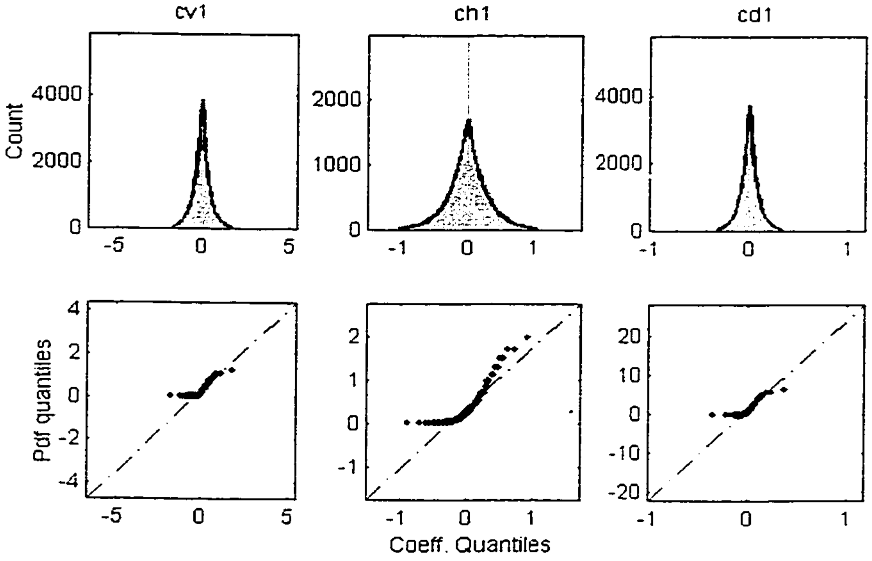


Figure 7-55: The Lapacien pdf versus the distribution of the thresholded details wavelet coefficients at the first level of decomposition.

straighter the line is, the closer to the assumed distribution the samples are. For a general α, the line is of slope $\frac{1}{\alpha}$. The statistics of the thresholded detail coefficients using the *SureShrink* operator are displayed in figures 7-55, 7-56, and 7-57. The three statistical tools, namely the bar plot, the Q–Q plot and the histograms show that the thresholded wavelet coefficients distribution can be fairly be modeled with the *Lapacien* pdf.

Global thresholding

The global thresholding compression based on the *VisuShrink* operator is first considered using both a soft and a hard thresholding rule. Figure 7-58 shows the compression results using a hard-thresholding rule with a universal threshold $\lambda = 4.9953$. The compression ratio is 93 : 1, with 12% of the original energy being discarded. The effect of the com-

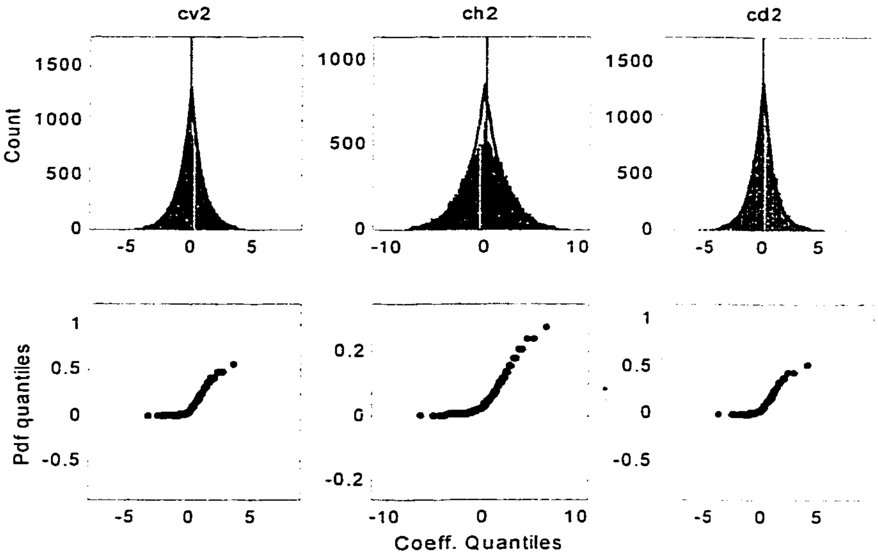


Figure 7-56: The Lapacien pdf versus the distribution of the thresholded details wavelet coefficients at the second level of decomposition.

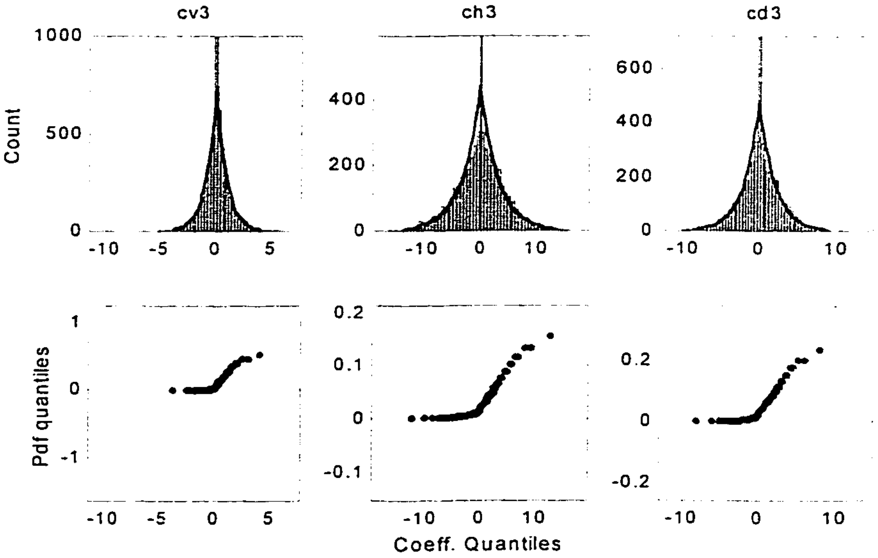


Figure 7-57: The Lapacien pdf versus the distribution of the thresholded details wavelet coefficients at the third level of decomposition.

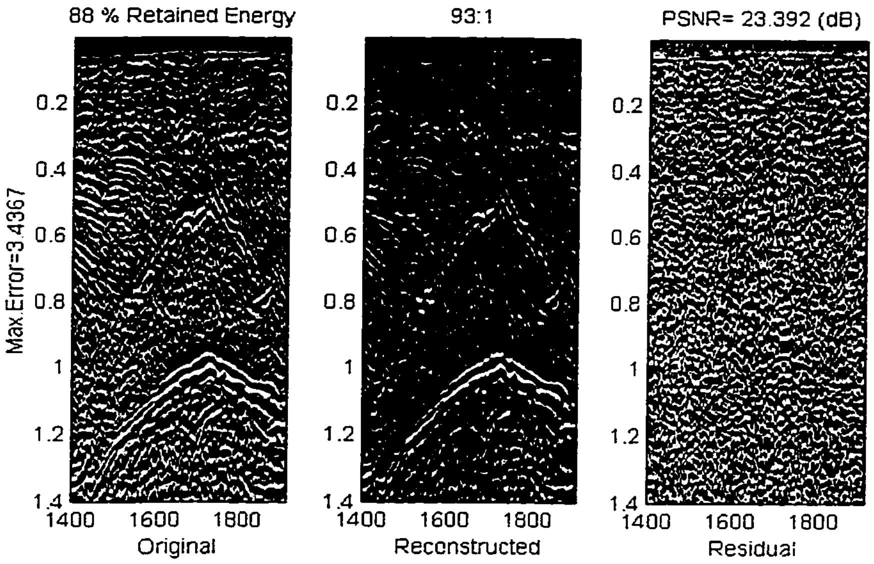


Figure 7-58: Compression results using VisuShrink hard-thresholding with a universal threshold $\lambda = 4.9953$ and a CDF(3,9) wavelet system using a VDD mode.

pression on the seismic traces is shown in figure 7-59 using a VAD mode. The L^{∞}-norm is 3.4367 and the difference section and its energy exhibit a substantial correlation along the horizontal direction as shown by the ACF of figure 7-60.

The compression performance using the soft thresholding operator with the same universal threshold are illustrated by the results of figures 7-61, 7-62 and 7-63. Notice that for the same compression ratio, 54% of data energy is discarded resulting in an important decrease in the PSNR and an increase in the correlation of the coefficients in the residual section. The latter exhibits some visible structures consisting principally of dipping events as shown in figure 7-62. An explantation for this performance deterioration could be the large value of the universal threshold, by which the thresholded coefficients are shrunk to the origin using a soft-thresholding operator. This causes an over-smoothing of the

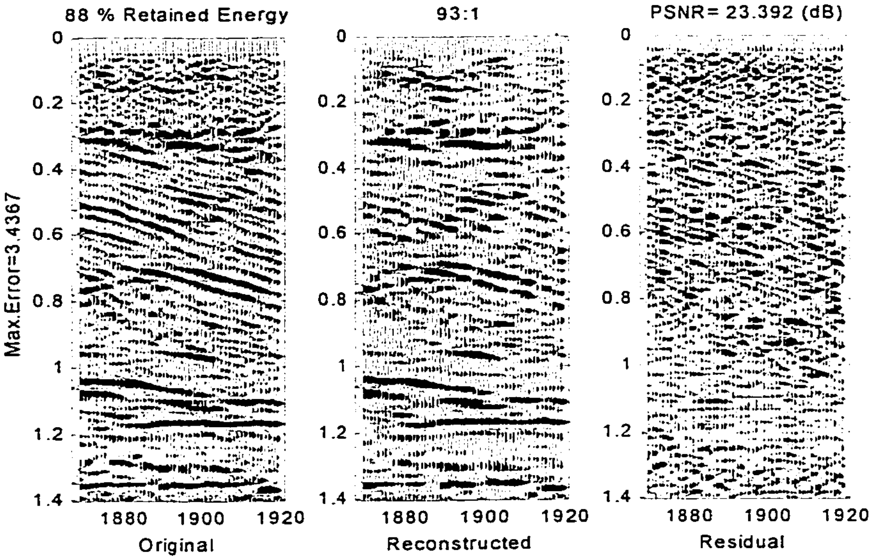


Figure 7-59: Compression results using VisuShrink hard-thresholding with a universal threshold $\lambda = 4.9953$ and a CDF(3,9) wavelet system using a VAD mode.

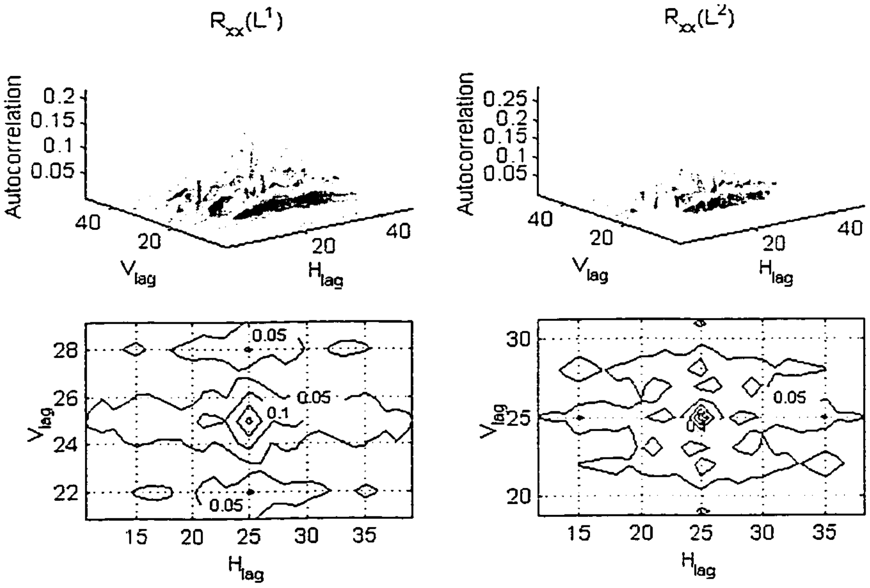


Figure 7-60: ACF of the L^1 and L^2-norms of the compression error using a VisuShrink hard-threshlod and the CDF(3,9) system. The L^∞-norm is 3.4367.

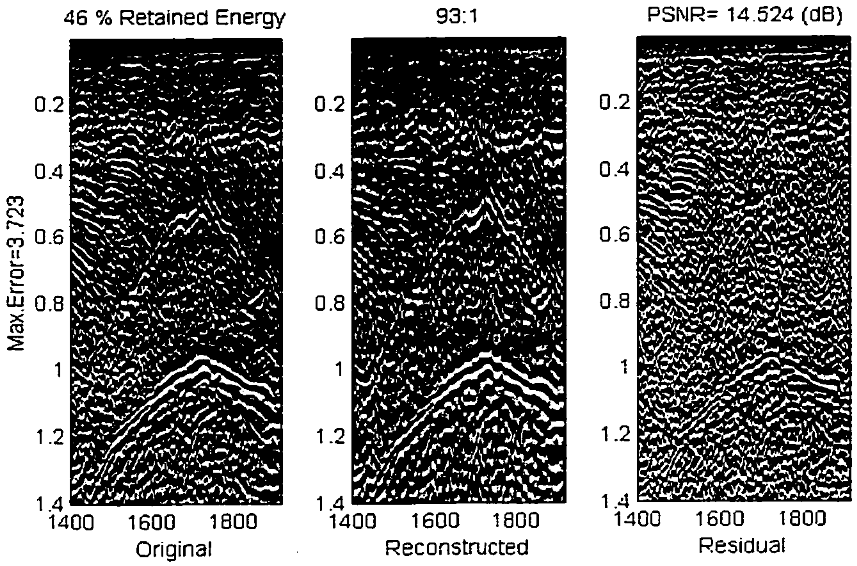


Figure 7-61: Compression results using VisuShrink soft-thresholding with a universal threshold $\lambda = 4.9953$ and a CDF(3,9) wavelet system uisng a VDD mode.

reconstructed data.

In order to study the impact of the wavelet bases choice on the compression performance, the *VisuShrink* compression is applied to the wavelet decomposition obtained using the FBI system. As the energy compaction power of the FBI bases is inferior to the one of the CDF(3,9) bases, the compression performance are poor as demonstrated by the results of figures 7-64, 7-65 and 7-66.

To end this comparative study, the DCT-based compression using the DCT-IV algorithm and 8x8 blocks is considered. In figure 7-67, the compression outcome is compared to the *VisuShrink*-based compression for a ratio of 93:1. Besides the poor PSNR, the zoomed view clearly shows the block artifact of the DCT-based compression.

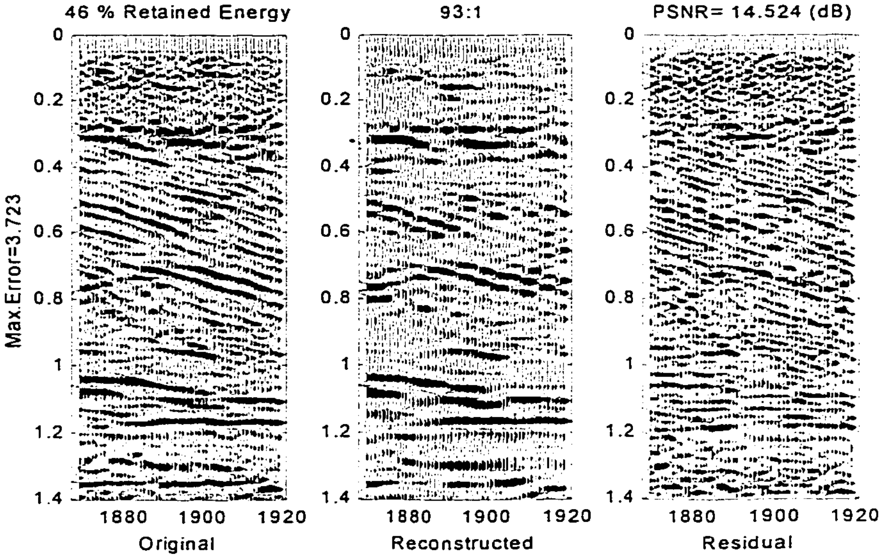


Figure 7-62: Compression results using VisuShrink soft-thresholding with a universal threshold $\lambda = 4.9953$ and a CDF(3,9) wavelet system uisng a VAD mode.

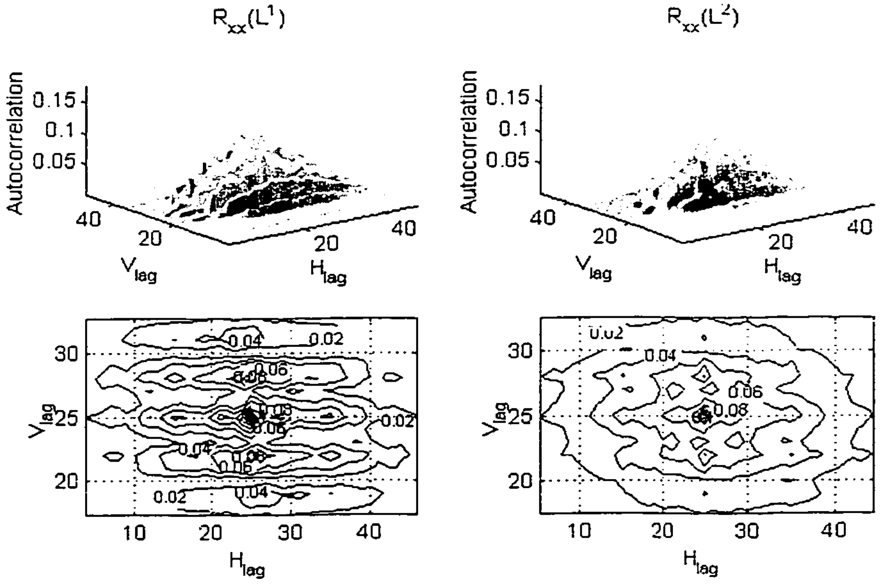


Figure 7-63: ACF of the L^1 and L^2-norms of the compression error using a VisuShrink soft-threshlod and the CDF(3,9) system. The L^∞-norm is 3.723.

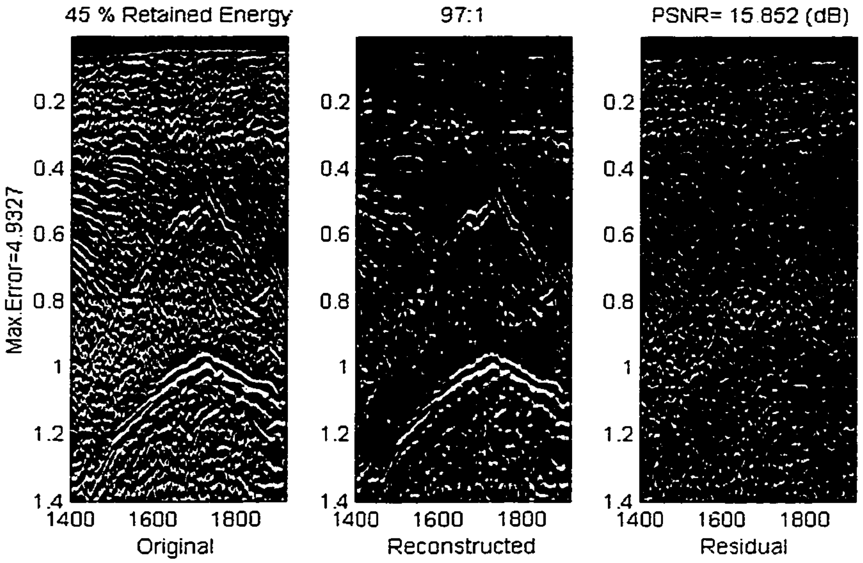


Figure 7-64: Compression results using VisuShrink hard-thresholding with a universal threshold $\lambda = 4.9953$ and a CDF(4,4) wavelet system using a VDD mode.

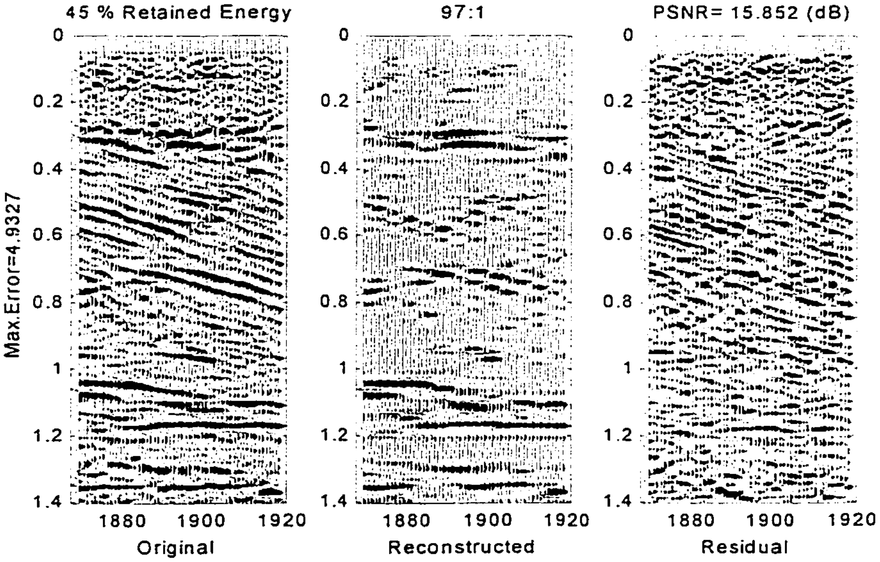


Figure 7-65: Compression results using VisuShrink hard-thresholding with a universal threshold $\lambda = 4.9953$ and a CDF(4,4) wavelet system using a VAD mode.

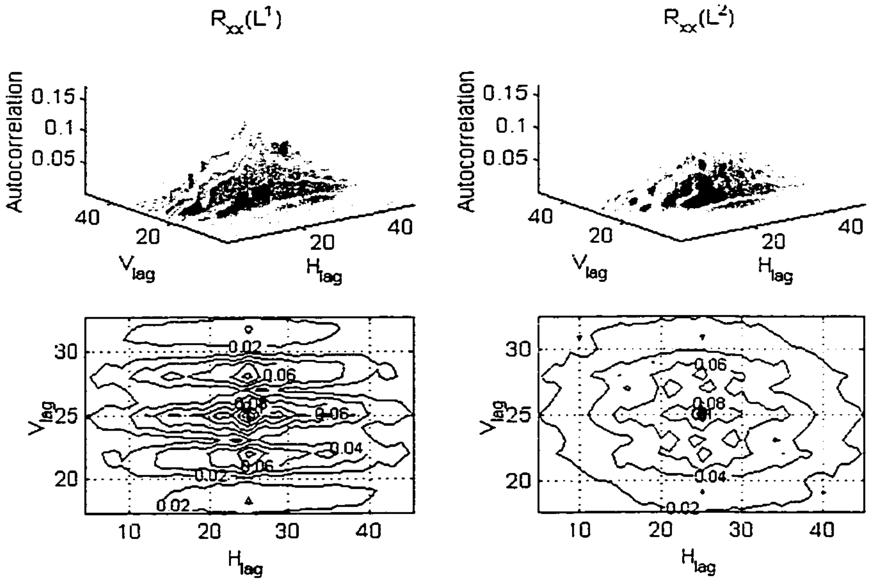


Figure 7-66: ACF of the L^1 and L^2-norms of the compression error using a VisuShrink hard-threshlod and the CDF(4,4) system. The L^∞-norm is 4.9327.

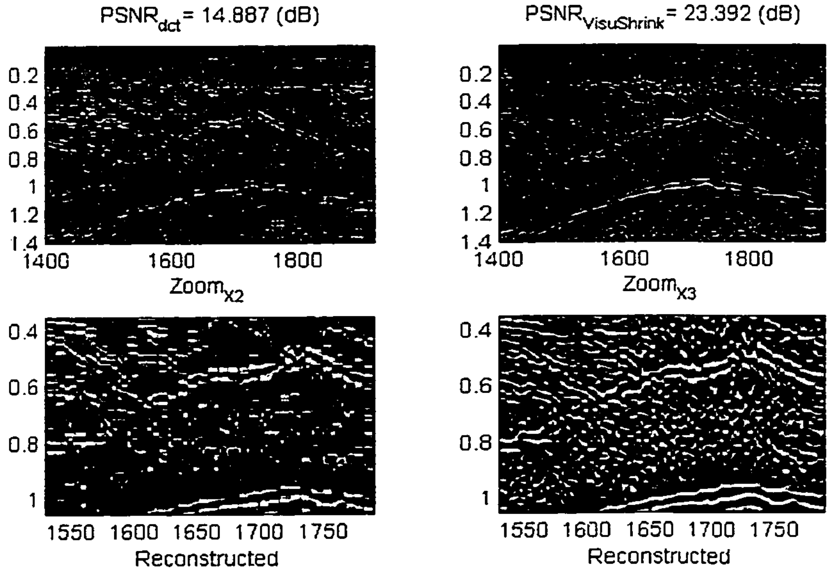


Figure 7-67: DCT versus VisuShrink compression for a ratio of 93:1.

	H	**V**	**D**
Level 1	*1.4222*	*3.1755*	*0.3534*
Level 2	*0.2338*	*0.2658*	*0.6446*
Level 3	*0.1928*	*0.3344*	*0.1975*

Table 7.1: Level- and orientation thresholds obtained using SURE principle.

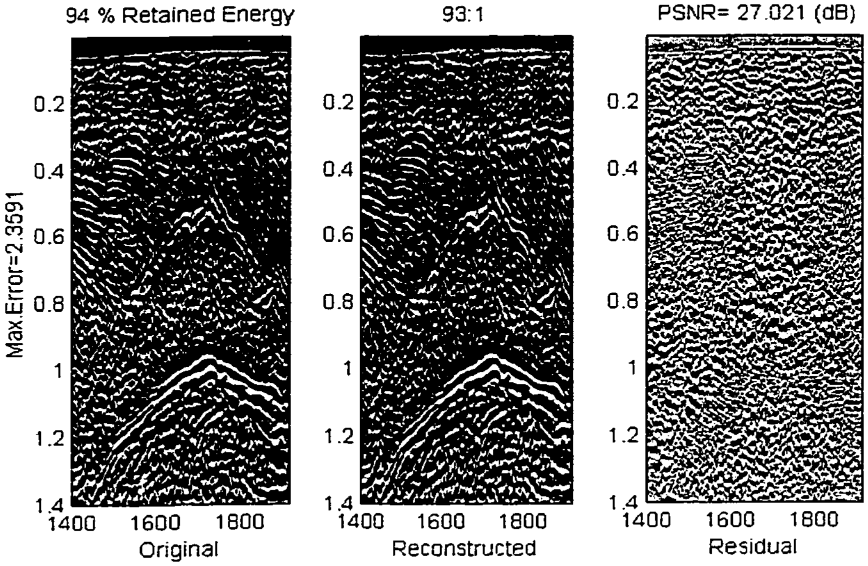


Figure 7-68: Compression results using SureShrink with the CDF(3,9) wavelet system using a VDD mode.

Adaptive thresholding

The last part of these experiments deals with the performance of the adaptive thresholding compression. We propose two different techniques. The first one is based on the *SureShrink* operator. The level-dependent thresholds are summarized in Table 7.1 and the compression results are presented in figures 7-68, 7-69, and 7-70. When compared to the *VisuShrink* compression, only 6% of the total energy is lost for the same compression ratio. The PSNR is higher, the magnitude of the error is smaller and the autocorrelation function of the error is well peaked, i.e. close to a δ-function. .

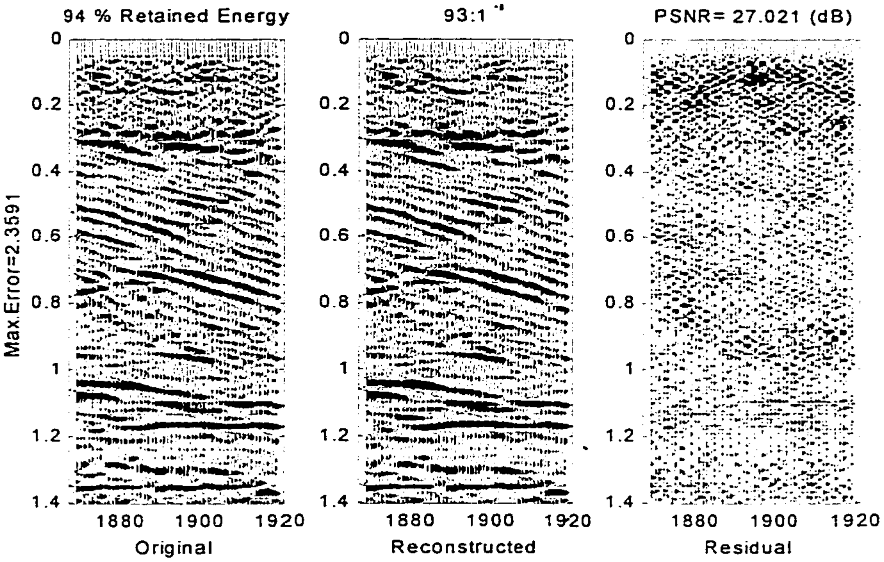


Figure 7-69: Compression results using SureShrink with the CDF(3,9) wavelet system using a VAD mode.

Finally, the compression results based on the *Birgé-Massarat* strategy are presented in figures 7-71, 7-72, and 7-73. It can be inferred that the performance of the latter is close to the one of the *SureShrink* compression.

7.4 Coherent Noise Removal

Ground roll is a surface wave energy that propagates along and near the surface with relatively low velocity, often with low frequency and usually with high amplitudes relative to other events of interest in land seismic surveys. The conventional wave-number (f-k) filtering which is based on the 2-D *Fourier* transform is not suitable to remove such a time- and space-variant correlated noise. An adaptive 2-D wavelet-based filtering technique is proposed. The rationale of the denoising strategy is to first generate a tree multiresolu-

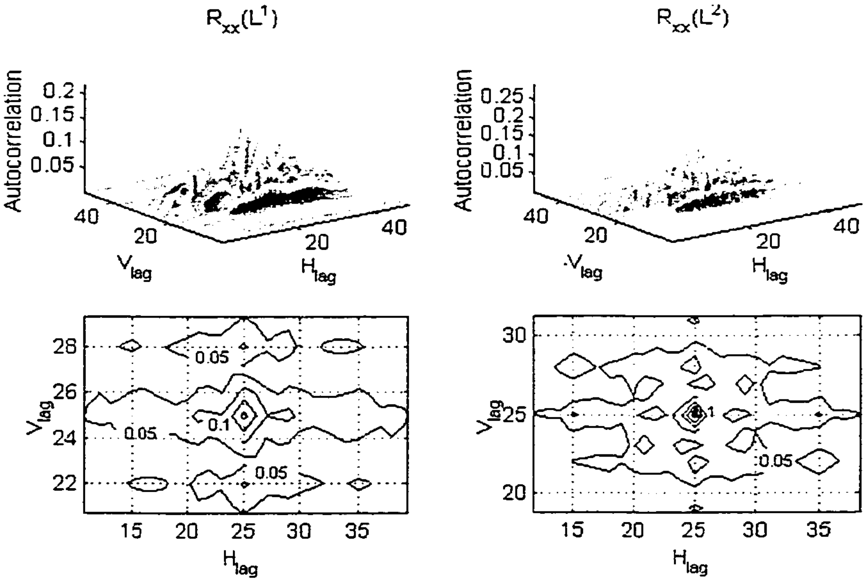


Figure 7-70: ACF of the L^1 and L^2-norms of the compression error using a SureShrink hard-threshlod and the CDF(3,9) system. The L^∞-norm is 2.3591.

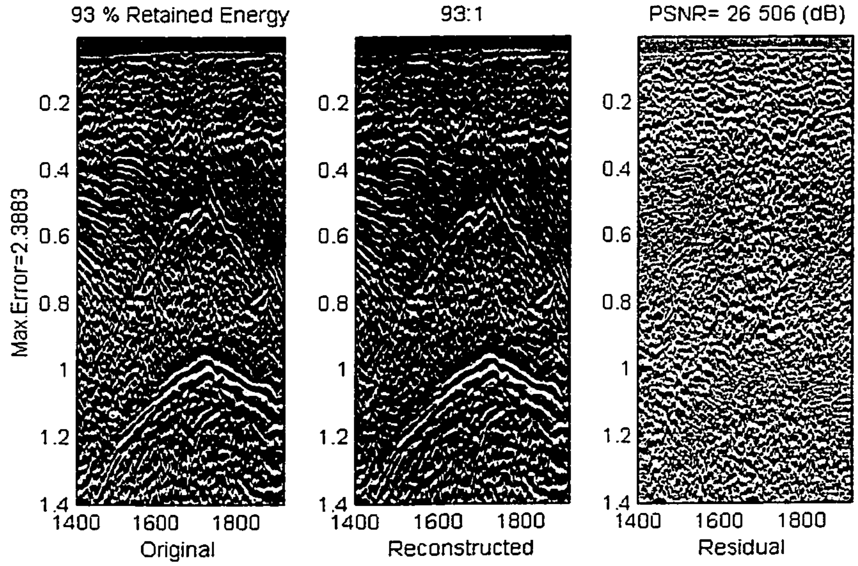


Figure 7-71: Compression results using Birgé-Massart shrinkage operator with the CDF(3,9) wavelet system using a VDD mode.

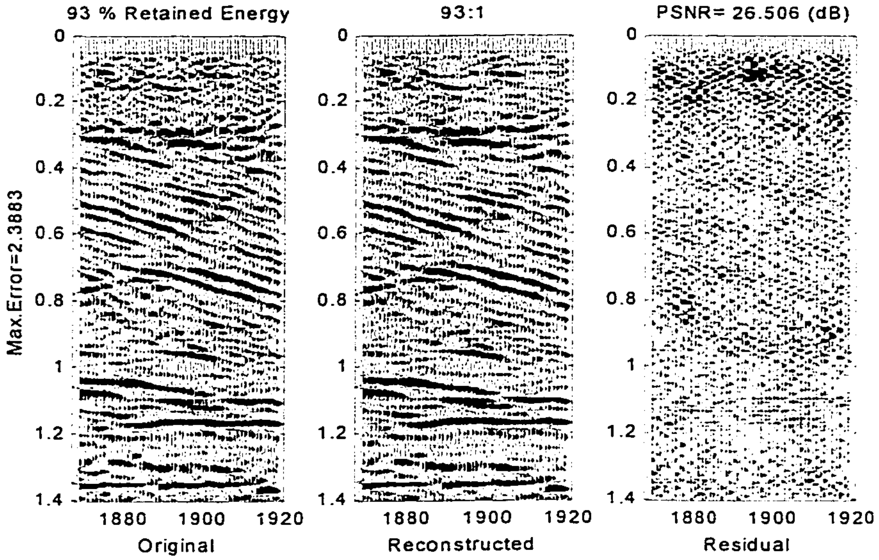


Figure 7-72: Compression results using Birgé-Massart shrinkage operator with the CDF(3,9) wavelet system using a VAD mode.

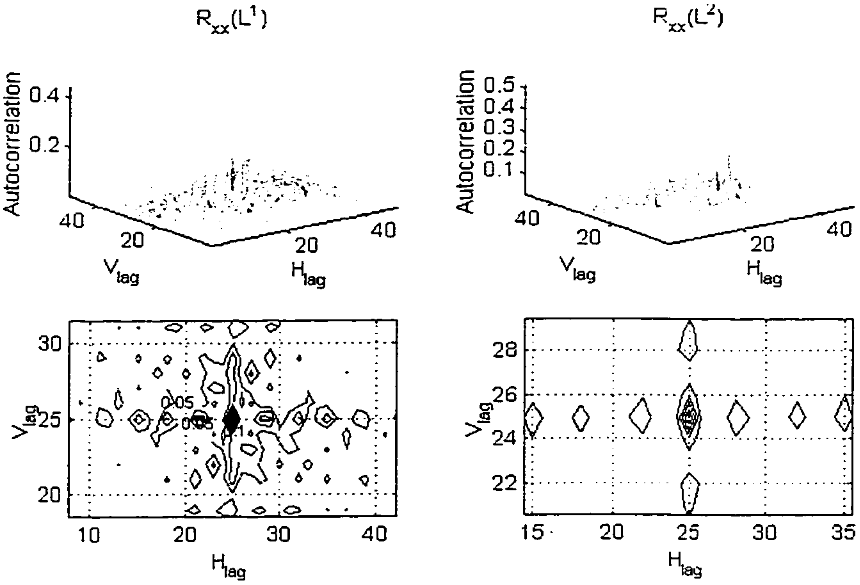


Figure 7-73: ACF of the L^1 and L^2-norms of the compression error using a SureShrink hard-threshlod and the CDF(3,9) system. The L^∞-norm is 2.3883.

tion decomposition of the data using a biorthogonal 2-D DWT. Then, the characteristic patterns of the mapping of the ground roll into the horizontal, vertical and diagonal detail subbands are analyzed. Next, the level-dependent thresholds are determined using the *SureShrink* technique. These are applied to all detail subbands but not the vertical subbands. Indeed, the 2-D vertical wavelet extracts most of the ground roll together with the refracted events. Consequently, the vertical thresholds are tuned to remove the former without seriously affecting the refractions. The denoised data is then reconstructed by the inverse wavelet transform.

The proposed technique is demonstrated with a seismic offset-time record from *Alaska* [109]. The seismic section shown in figure 7-74, exhibits a series of reflections, refractions and is corrupted by random noise and ground roll. A CDF(3,9) biorthogonal wavelet filter pair is used and a three-level multiresolution decomposition is generated. From figure 7-75, which depicts the reconstructed vertical subband at level 1, one can see that most of the ground roll energy is confined to that subband. An explanation to this could be that the slope of the envelope containing the ground roll lies below the diagonal, whereas the reflections lie above it. However, the refracted events share the same vertical subband and the relative threshold is consequently tuned to remove the latter only.

Figure 7-76 shows the denoised data after reconstruction. Notice that the undesirable ground roll is significantly filtered with a little effect on the rest of the record while the random noise is almost killed. Finally, the dynamic of the reconstructed data is honored by applying an *automatic gain control* (AGC) to restore the data above and below the removed ground roll.

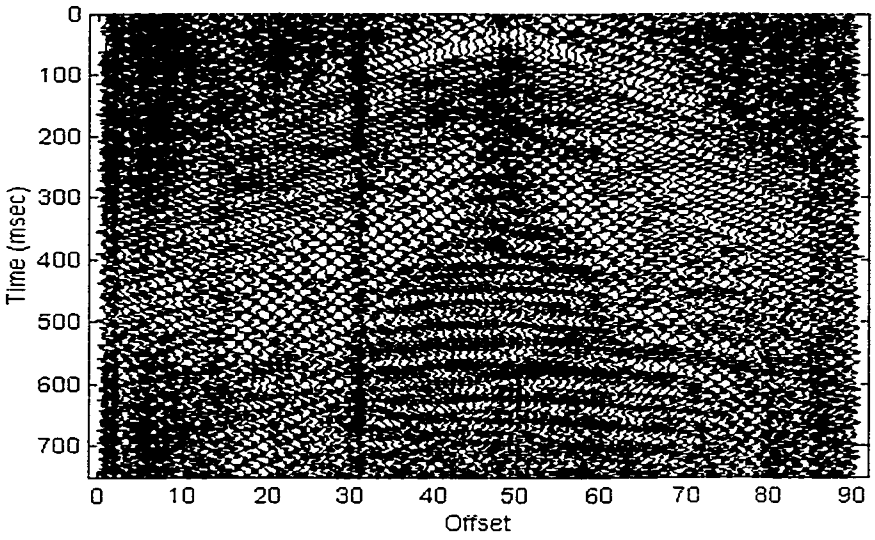


Figure 7-74: Original section corrupted by ground roll and random noise.

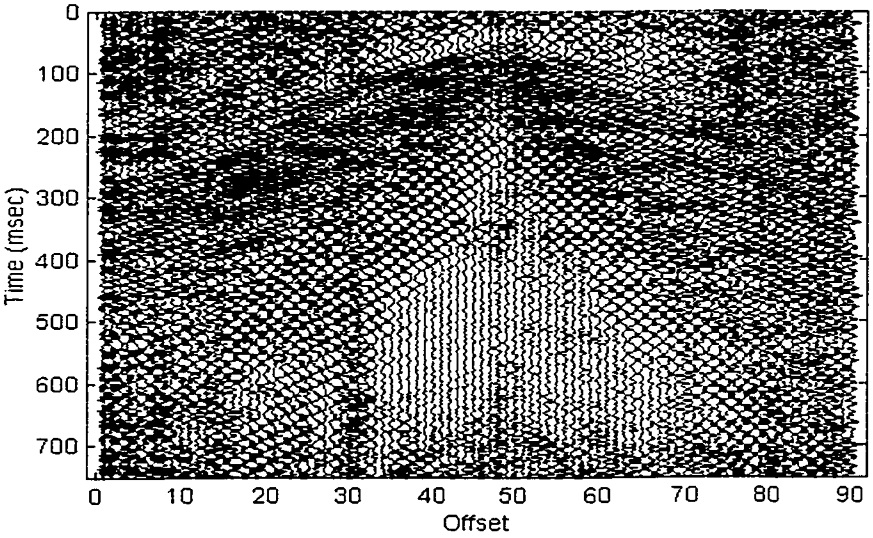


Figure 7-75: Reconstructed vertical subband@L1

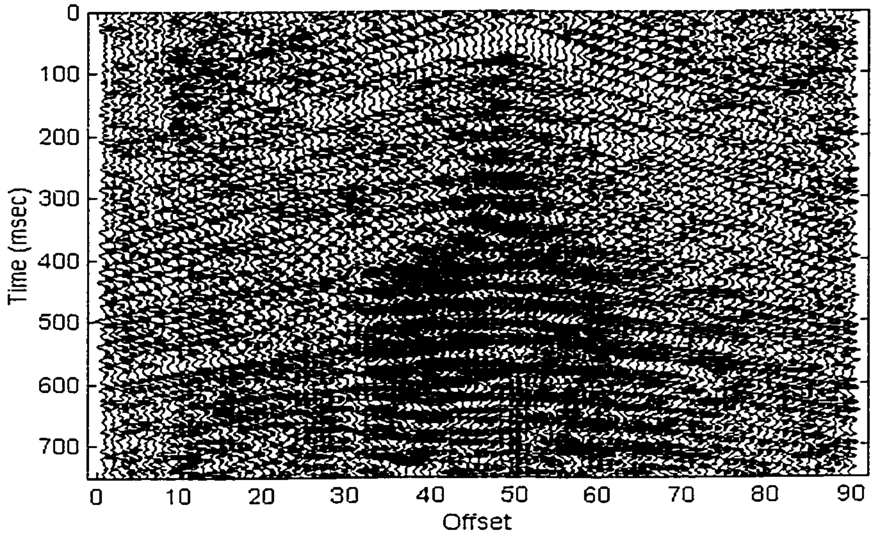


Figure 7-76: Denoised section.

Chapter 8

CONCLUSIONS AND RECOMMENDATIONS

The primary objective of this dissertation is a contribution to the development of modern and efficient compression techniques to solve the problems faced by the oil & gas industry in managing ever increasing data sets.

The proposed compression scheme consists of a 2-D fast wavelet transform (FWT), an adaptive quantizer and an entropy encoder. Both orthogonal and biorthogonal wavelet systems have been used. The fast wavelet transform is implemented using both the lattice structure and the lifting scheme. However, the lifting scheme is more versatile and has many advantages. Firstly, it is a very powerful technique to factorize every unitary and non-unitary FIR filter bank into a finite sequence of simple cells known as *ladder structures*, reducing the complexity by half, i.e. $\mathcal{O}(N/2)$. Then it allows for an in-place implementation of the FWT, saving thus any auxiliary memory usage, which turns out to

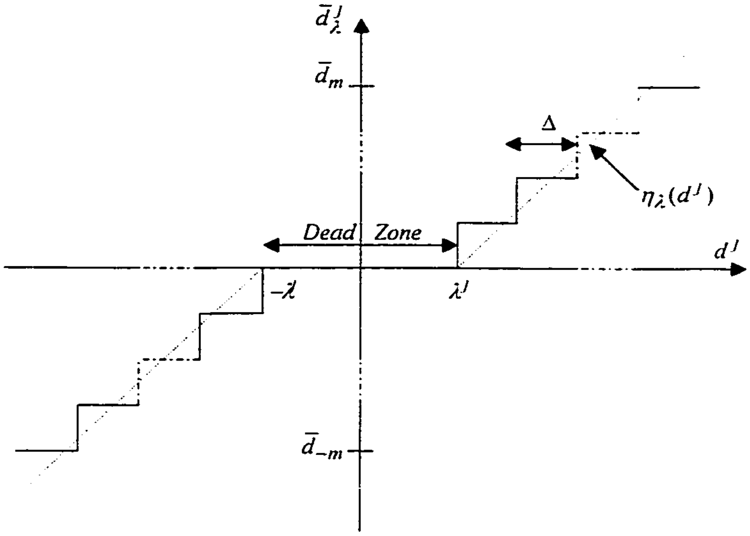


Figure 8-1: The adaptive *TUSQ* constitues the compression engine. The parameters λ, m and $\triangle$ are determined via wavelet shrinkage and *MDL*, respectively.

be of the outmost importance when compressing large seismic data sets. Secondly, every lifted DWT is immediately invertible with exactly the same complexity as for the forward transform. Lastly, the lifting scheme exhibits an inherent parallelism that can be exploited if multiprocessors workstations are used to perform the on-line compression/decompression of seismic data.

The adaptive compression engine is depicted in figure 8-1. It involves techniques from non-linear approximation and information theories, namely the *wavelet shrinkage* [29] and the *MDL* [79]. The scale and direction-dependent dead-zone or threshold λ^j is determined for each detail subband using the *ShureShrink* principle [28],[30] and the *Birgé-Massart* strategy [6]. The bit allocation problem, which consists in finding the optimal number of bins m and the bin-size Δ, has been tackled in the sense of *Kolmogorov* [64] by using the *MDL* principle.

A thorough experimental study has been performed using a real seismic data set provided by *Saudi ARAMCO* and a multitude of issues in seismic compression have been explored. Firstly, the mechanisms of the seismic wavelet expansion have been investigated. It has been demonstrated that it is preferable to use filters with a small support width to avoid the important increase in the dynamical range due to overlapping singularities. However, it is judicious to seek a trade-off between the smoothness of the wavelet bases and their support. Indeed, it has been ascertained that an excessive increase in the number of vanishing moments (VM) yields only a marginal improvement in the performance of the wavelet bases. Nevertheless, the number of VM plays an important role in compacting the energy and reducing entropy of the input data. For that reason, the daublet-2 wavelet system that is endowed with only two VM, is less efficient than the daublet-8 system.

Secondly, different classes of wavelet systems have been tested. It has been shown that the biorthogonal system offers more flexibility than the orthogonal one. In fact, by using different wavelet filters pairs for the analysis and synthesis stages, the performance of the wavelet expansion and the subsequent compression can be dramatically improved. Furthermore, the performance of two members of the B-splines biorthogonal family, namely the CDF(3,9) and the CDF(4,4) have been compared. The CDF(3,9) wavelet system demonstrated a remarkable performance in either compacting the energy and smoothing the reconstructed data, when the shorter filter is used in the analysis stage and the longer one is used in the synthesis stage. The CDF(3,9) performed better than the CDF(4,4) wavelet system. Although the latter has been successfully applied for compressing the fingerprints database of the FBI, the experimental results clearly show that it performs poorly as compared to the CDF(3,9) system.

Thirdly, various compression techniques have been applied and the results have been compared using various metrics and compression scores. For the global thresholding compression, the *VisuShrink* operator has been used with two options: a hard- and a soft-thresholding rules. For the sake of comparison, the DCT compression has also been considered using the same global threshold. The *VisuShrink* compression performs better when the hard-thresholding rule is used. This is because the universal threshold, which is a function of the data size, is quite large and that using it with the soft-thresholding will oversmooth the reconstructed data. The DCT compression performs as poorly as the *VisuShrink* with a soft thresholding rule in addition to the block effect. The proposed adaptive thresholding method used two shrinkage operators, namely the *SureShrink* and the *Birgé-Massart*. It has been demonstrated that the adaptive thresholding is superior to the global thresholding. Elsewhere, it has been found that the *SureShrink* compression is slightly more efficient than the *Birgé-Massart* compression. Finally, the ability of the *SureShrink* to successfully filter the ground roll in the 2-D wavelet domain has been studied. It can be concluded that based on the SURE principle and on the characteristic patterns of the ground roll, an adaptive nonlinear soft thresholding 2-D filter is able to attenuate both the ground roll and the random noise without seriously affecting the rest of the record.

The chief contributions are summarized as follows:

- Introduction of the lifting scheme for seismic compression

- Combined optimization of the compression stages
- Design of an adaptive quantizer based on the wavelet shrinkage and the MDL principle
- Simultaneous compression and denoising of seismic data
- Filtering of the ground roll in the wavelet domain

These contributions as well as related materials have been published in [1],[55],[56],[57],[58],[59],[60],[61].

Based on experiences and insights acquired during this endeavor, the following recommendations are made for future investigations:

- The impact of seismic compression of the interpretability of the final product
- Seismic volume compression using the 3-D lifted wavelet transform
- Exploitation of the inherent parallelism of the lifting scheme for performing interactive visualization based on volume rendering techniques.

Bibliography

[1] Abdul-Jauwad, S. H. and Khène, M. F., 2000, *Two-dimensional wavelet-based ground roll filtering.* 70th International Meeting, Society of Geophysical Exploration (SEG). Calgary, Canada, August 6-11.

[2] Abramovich, E., Sapatinas, T., and Silverman, B., 1998, *Wavelet thresholding via a Bayesian approach.* J. Royal Statist. Soc., Series B, 60:725-749.

[3] Barry, K., Cavers, D., and Kneale, C., 1975, *Recommended standards for digital tape format.* Geophysics 40(2):344-352.

[4] Berger, T., 1971, *Rate distortion theory.* Prentice-Hall.

[5] Bernhardt, T. and Peacock, J., 1976, *Encoding techniques for the vibroseis system.* Geophysics, 56, 1058:1060.

[6] Birgé, L. and Massart, P, 1997, *From model selection to adaptive estimation.* In Festschrift for Lucien Le Cam: *Research Papers in Probability and Statistics* Pollard D., Torgersen E. and Yang G. (Eds.), 55-87. Springer-Verlag, New York.

[7] Bordley, T., 1983, *Linear predictive coding of marine seismic data.* IEEE trans. on Acoust. Speech and Signal Processing, 31(4):828-835.

[8] Bosman, C. and Reiter, E., 1993, *Seismic data compression using wavelet transforms.* 63rd Annual Internat. Mtg., SEG, Expanded Abstracts, 1261-1264

[9] Bradley, J., Brislawn, C., and Hopper, T., 1993, *The FBI wavelet/scalar quantization standard for gray-scale fingerprint image compression.* Proc. SPIE 1961.

[10] Bradley, J. Fei, T., and Hildebrand, S., 1996, *Wavelet compression for 3-D depth migration.* 66th Annual Internat. Mtg., SEG, Expanded Abstracts, 1627-1629.

[11] Brown, A., 1996, *Interpretation of three-dimensional seismic data.* AAPG Memoir 42, 4th Edition.

[12] Burrus, C., Gopinath R., and Guo, H., 1998, *Introduction to wavelets and wavelet transforms.* Prentice Hall.

[13] Burt, P., and Adelson, E., 1983, *The Laplacien pyramid as a compact image coder.* IEEE trans. on Communication, 31(4):532-540.

[14] Chang G., Yu B. and Vetterli M., *Adaptive wavelet thresholding for image denoising and Compression.* http://cm.bell-labs.com/cm/ms/departments/sia/binyu

[15] Chen, T., 1995, *Seismic data compression.* Msc. thesis, CWP-196, Colorado School of Mines.

[16] Clark, R., 1996, *Digital compression of still images and video.* Academic Press.

[17] Cohen A., Daubechies I., Feauveau J., 1992, *Biorthogonal bases of compactly supported wavelets.* Commun. on Pure Appl. Math., 45:485-560.

[18] Cohen, A., Daubechies, I., Jawreth, B., and Vial, P., 1993, *Multiresolution analysis, wavelets and fast wavelet transform on an interval.* CRAS Paris, A(316):417-421.

[19] Daubechies, I., 1988, *Orthonormal bases of compactly supported wavelets.* Comm. Pure Appl. Math., 41:909-996.

[20] Daubechies, I., 1992, *Ten lectures on wavelets.* SIAM, Philadelphia.

[21] Daubechies, I and Lagarias, J., 1992, *Two-scale difference equations, II.* SIAM J. Math. Anal., 22:1031-1079.

[22] Daubechies, I., and Sweldens, W., 1998, *Factoring wavelet transforms into lifting steps.* J. Fourier Anal. Appl., 4:245-267.

[23] Dean, D. and Mosher, C., 1996, *Effect of seismic data compression on interpretation projects: An AVO example.* 66th Annual Internat. Mtg., SEG, Expanded Abstracts, 2043.

[24] DeBoor, C., Höllig, K., and Riemenschneider, S., 1993, *Box splines.* Springer Verlag.

[25] DeVore, R., Jawreth, B., and Lucier, B., 1992, *Image compression through wavelet transform coding.* IEEE Trans. Information Theory 38(2):719-746.

[26] Dessing, F., 1997, *A wavelet transform approach to seismic processing.* PhD thesis TUD.

[27] Diller, D., Hale, D. and Foy, R., 1996, *Seismic compression exchange standardization via the self-extracting archive model.* 66th Annual Internat. Mtg., SEG, Expanded Abstracts, 2039-2040.

[28] Donoho, D. and Johnstone, I., 1994, *Ideal spatial adaptation by wavelet shrinkage.* Biometrika, 81:425-455.

[29] Donoho D., Johnstone I., Kerkyacharian G. and Picard D., 1995, *Wavelet shrinkage: asymptotia?* J. of the Royal Statistical Society, Series B, 57(2):301-369.

[30] Donoho, D. and Johnstone, I., 1995, *Adapting to unknown smoothness via wavelet shrinkage.* J. of the Amer. Statistical Assoc. 90:1200-1224.

[31] Donoho, D., 1995, *Denoising by soft thresholding.* IEEE Trans. Information Theory, 41(3): 613-627.

[32] Donoho, P., Ergas, R., and Villasenor, J., 1995, *High-performance seismic trace compression.* 65th Annual Internat. Mtg., SEG, Expanded Abstracts, 160-163.

[33] Donoho, P., Ergas, R., Polzer, R. and Villasenor, J., 1996, *Compression optimization by multidimensional wavelet transforms and data dependent quantization.* 66th Annual Internat. Mtg., SEG, Expanded Abstracts, 2042.

[34] Donoho, P., Ergas, R., and Polzer, R., 1999, *Development of seismic data compression diagnostics.* 61st Annual Mtg., EAGE, Session P104.

[35] Donoho, P., Ergas, R., and Polzer, R., 1999, *Development of seismic data compression methods for reliable low-noise performance.* 69th Annual Internat. Mtg., SEG, Expanded Abstracts, 236-239.

[36] Ergas, R., Polzer, R., Donoho, P., and Galibert, P-Y, 1996, *Pitfalls in compressing land seismic trace data.* 58th Annual Mtg., EAGE, Extended Abstracts, Session P156.

[37] Ergas, R., 1996, *Seismic data compression- A key technology for the future.* 66th Annual Internat. Mtg., SEG, Expanded Abstracts, 1904.

[38] Ergas, R., Polzer, R., Donoho, P., and Villasenor, J., 1996, *Measuring seismic data compression: What losses are acceptable?* 66th Annual Internat. Mtg., SEG, Expanded Abstracts, 2041.

[39] Esteban, D. and Galand, C., 1977, *Application of quadrature mirror filters to split-band voice coding scheme.* Int. Conf. ASSP, 191-195, Hardtford, Connecticut.

[40] Goupillaud, P., 1976, *Signal design in the Vibroseis technique.* Geophysics, 41, 1921:1304.

[41] Gresho, A. and Gray, R., 1992, *Vector quantization and signal compression.* Kluwer Academic Publishers.

[42] Guo, H. and Burrus, C., 1996, *Phase-preserving compression of seismic data using self-adjusting wavelet transform.* NASA Combined Industry, Space and Earth Science Data Compression Workshop, Snowbird.

[43] Hall, M., Monk, D., and Reiter, E., 1995, *An evaluation of seismic data compression on the interpretability of the final product.* 57th Annual Mtg., EAGE, Session B 35.

[44] Hardage B., Levey R., Pendleton V., Simmons J., and Edson R., 1994, *A 3-D seismic history evaluation fluvially deposited thin-bed reservoirs in a gas-producing property.* Geophysics 59(11): 1650-1665.

[45] Helle, H. and Daehlen, M., 1991, *Compression of data for modeling geological and geophysical surfaces - A new approach based on B-splines.* 61st Annual Internat. Mtg., SEG, Expanded Abstracts, 328-331.

[46] Hewlett, C. and Hatton, L., 1995, *Seismic data compression with CD-ROM archiving.* 57th Annual Mtg., EAGE, Session: B034.

[47] Holschneider, M., 1995, *Wavelets: An analysis tool.* Clarendon Press, Oxford.

[48] Hutchins, A. and Anderson, R. (Eds.), 1996, *4-D seismic handbook.* World Oil, Gulf Publishing Co., Houston.

[49] Jain , A., 1989, *Fundamentals of digital image processing.* Prentice Hall.

[50] Jayant, N. and Noll, P., 1984, *Digital coding of waveforms - Principles and applications to speech and video.* Prentice Hall, Englewood Cliffs.

[51] Johnstone, J., 1980, *A filter family designed for use in quadrature mirror filter banks.* ICASSP, 291-294, Denver, USA.

[52] JPEG2000, Requirements Ad hoc group, July 1999, *JPEG2000 requirements and profiles version 6.0.* WG1 Vancouver Mtg.

[53] Kaiser, G., 1994, *A friendly guide to wavelets.* Birkhäuser.

[54] Kalker, T., and Shah, I., 1992, *Ladder structures for multidimensional linear phase perfect reconstruction filter banks and wavelets.* pp 12-20, Proc. SPIE 1818.

[55] Khène, M. F., and Abdul-Jauwad, S. H., 1998, *The wavelet transform: A potential tool for seismic data compression.* 5th Mtg. of Saudi Society of Earth Sci. 29-29 Oct. KFUPM.

[56] Khène, M. F., and Abdul-Jauwad, S. H., 1999, *Compression of seismic images with the discrete wavelet transform.* Geovision99, International Symposium on Imaging Applications in Geology, Liège (Belgium), 141-144.

[57] Khène, M. F., and Abdul-Jauwad, S. H., 2000, *Adaptive wavelet-based seismic compression.* GeoArabia, 5(1): 126-127.

[58] Khène, M. F., and Abdul-Jauwad, S. H., 2000, *Efficient seismic volume compression using the lifting scheme.* SPIE 45th Annual Mtg. July 30-August 4, San Diego.

[59] Khène, M. F., and Abdul-Jauwad, S. H., 2000, *Adaptive Seismic Compression by Wavelet Shrinkage.* SAAP2000, 10th IEEE Workshop on Statistical Signal and Array Processing, August 14-16, Pocono Manor, Pennsylvania.

[60] Khène, M. F., and Abdul-Jauwad, S. H., 2000, *Efficient seismic compression using the lifting scheme.* 70th International Meeting, Society of Geophysical Exploration (SEG). Calgary, Canada, August 6-11.

[61] Khène, M. F., and Abdul-Jauwad, S. H., 2000, *Seismic signals compression using wavelet shrinkage and minimum description length.* 5th IMA International Conference on Mathematics in Signal Processing, Warwick, UK, December 2000.

[62] Kou, W., 1995, *Digital image compression: Algorithms and standards.* Kluwer Academic Publisher.

[63] Kromander, T., 1989, *Some aspects of perception based image coding.* Ph.D. Thesis, No. 203, Linköping University, Sweden.

[64] Li M. and Vitany P., 1997, *An introduction to Kolmogorov complexity and its applications.* Springer-Verlag.

[65] Liu, J., and Moulin, P., 1997, *Complexity-regularized image denoising.* Proc. IEEE int. Conf. Image Processing, vol. II:370:373.

[66] Liner, C., 1999, *Elements of 3-D seismology.* Pennwell.

[67] Luo, Y. and Schuster, G., 1992, *Wave packet transform and data compression:* 62nd Annual Internat. Mtg., SEG, Expanded Abstracts, 1187-1190.

[68] Mallat, S., 1998, *Wavelet tour of signal processing.* Academic Press.

[69] Marr,D., 1982, *Vision.* Freeman and Company.

[70] Meyer, Y., 1993, *Wavelets: Algorithms and applications.* SIAM.

[71] Misitti M., Missiti Y., Oppenheim. G and Poggi J-M., 1997, *Matlab wavelet toolbox.* The MathWorks, Inc.

[72] Mosher, C., 1996, *Towards a SEG standard for data compression: A request for technology.* 66th Annual Internat. Mtg., SEG, Expanded Abstracts, 2045.

[73] Mougenot D. and Al-Shakhis A., 1998, *Depth imaging a pre-salt faulted block: A case study from the Midyan basin (Red Sea).* Saudi Aramco J. of Technology, 2-10.

[74] Mougenot, D., 1998, *Suivre le pétrole à la trace.* La Recherche, No. 312:34-38.

[75] Pritchett, W., 1990, *Acquiring better seismic data.* Chapman and Hall.

[76] Ramstad, T., Aase, S., and Husoy, J., 1995, *Subband compression of images: Principles and examples.* Elsevier.

[77] Reiter, E. and Heller, P., 1994, *Wavelet transform-based compression of NMO-corrected CDP gathers.* 64th Annual Internat. Mtg., SEG, Expanded Abstracts, 731-734.

[78] Reiter, E., 1996, *A quantitative comparison of 1, 2 and 3 dimensional wavelet compression methods for seismic data.* 66th Annual Internat. Mtg., SEG, Expanded Abstracts, 1630-1633.

[79] Rissanen J., 1989, *Stochastic complexity in statistical inquiry.* World Scientific, Singapore.

[80] Rioul, O., 1993, *A discrete-time multiresolution theory.* IEEE trans. on Signal Processing, 41(8):2591-2606.

[81] Rosenfeld, A. Ed., 1984, *Multiresolution techniques in computer vision.* Springer-Verlag.

[82] Rosten, T., Lervik, J., Ramstad, T., and Amundsen, L., 1996, *Subband compression of seismic stack sections.* 66th Annual Internat. Mtg., SEG, Expanded Abstracts, 1623-1626.

[83] Rosten, T., Lervik, J., Balasingham, I. and Ramstad, T. A., 1997, *On the optimality of filter banks in subband compression of seismic stack sections.* 67th Annual Internat. Mtg., SEG, Expanded Abstracts, 1338-1341.

[84] Saito N., 1994, *Simultaneous noise suppression and signal compression using a library of orthonormal bases and the MDL criterion.* Wavelets in Geophysics, Foufoula-Georgiou E. and Kumar E. (Eds.), Academic Press, 229:324.

[85] Sheriff, R., 1989, *Geophysical methods.* Prentice Hall, Englewood Cliffs.

[86] Singall, P. and Panagos, P., 1978, *Data compression in microprocessor-based data acquisition systems.* IEEE trans. Geoscience Electronics, 16(10):323-332.

[87] Spanias A., Johnsson, S., and Stearns, S., 1991, *Transform Methods for Seismic Data Compression.* IEEE Trans. Geoscience and Remote Sensing, 29(3): 407-416.

[88] Stein, C., 1981, *Estimation of the mean of a multivariate normal distribution. Annals of Statistics*, 9(6):1135-1151.

[89] Steinlein, O. and Hjelle, O., 1992, *Compact digital representation of seismic data.* 62nd Annual Intern Mtg. SEG, Extended Abstracts, 1176-1178.

[90] Stigant, J., Ergas, R., Donoho, P., Minchella, A. and Galibert, P-Y., 1995, *Field trial of seismic compression for real-time transmission.* 65th Annual Internat. Mtg., SEG, Expanded Abstracts, 960-962.

[91] Stollnitz, E., DeRose, T., and Salesin, D., 1996, *Wavelets for computer graphics: Theory and applications.* Morgan Kaufmann Publishers, Inc., California.

[92] Strang, G. and Nguyen, T., 1996, *Wavelets and filter banks.* Wellesley, Cambridge.

[93] Sullivan G., 1996, *Efficient scalar quantization of exponential and Laplacian random variables.* IEEE trans. on Information Theory, 42(5): 1365-1374.

[94] Sweldens, W., 1995, *The lifting scheme: A new philosophy in biorthogonal wavelet constructions.* In A. F. laine and M. Unser, editors, *Wavelet Applications in Signal and Image Processing* III, pp 68-79. Proc. SPIE 2569.

[95] Sweldens, W., 1996, *The lifting scheme: A custom-design construction of biorthogonal wavelets.* Appl. Comput. Harmon. Anal., 3(2), 186-200.

[96] Sweldens, W., 1997, *The lifting scheme: A construction of second generation wavelets.* SIAM J. Math. Anal., 29(2):511-546.

[97] The Leading Edge, 1998, *3-D seismic interpretation*, Special section, 17(9).

[98] The Leading Edge, 1999, *Land acquisition*, Special section, 18(10).

[99] The Leading Edge, 1999, *Multicomponent offshore*, Special section, 18(11).

[100] Vaidyanathan, P., 1993, *Multirate systems and filter banks.* Prentice Hall, Englewood Cliffs.

[101] Vassiliou, A. and Wickerhauser, V., 1997, *Comparison of wavelet images coding schemes for seismic data compression.* 67th Annual Internat. Mtg., SEG, Expanded Abstracts, 1334-1337.

[102] Vermeer, P., Bragstad, H. and Orr, C., 1996, *Aspects of seismic data compression.* 66th Annual Internat. Mtg., SEG, Expanded Abstracts, 2031-2034.

[103] Vermeer, P., 1999, *Compression of field data within system specifications.* 69th Annual Internat. Mtg., SEG., Expanded Abstracts.

[104] Vetterli, M. and Kovačević, J. 1995, *Wavelets and subband coding.* Prentice Hall PTR, New Jersey.

[105] Von Flatern, R., 1995, *Adding time makes seismic data a production tool.* Petroleum Engineer International, 12:17-21.

[106] Waldemar, P., Rosten, T. and Ramstad, T., 1997, *Compression of seismic stack sections using singular value decomposition.* 67th Annual Internat. Mtg., SEG, Expanded Abstracts, 1342-1345.

[107] Westernik, P., 1989, *Subband coding of images.* Ph.D. thesis. Technical University of Delft.

[108] Wood, L., 1974, *Seismic data compression methods.* Geophysics, 39(4): 499-525.

[109] Yilmaz, O., 1987, *Seismic Data Processing.* SEG, Tulsa, OK.

Vita

- Faouzi M. KHENE

- Born in Annaba (Algeria), on June 02, 1968

- Permanent Address:

  6, Rue Asfour Chérif, Beauséjour

  Annaba 23000

  Algeria

Email: mfkhene@hotmail.com

- Received Bachelor of Science (BSc.) degree in Electronics and Control from University of Annaba, Algeria in June 1990.

- Received Master of Science (MSc.) degree in Electrical Engineering from *Ecole Nationale Polytechnique*, Algeria in October 1993.

- Joined Ph.D program, in Electrical Engineering Department at KFUPM, Dhahran, Saudi Arabia, in September 1995.

- Research Areas: Wavelets, Image & Signal processing, Computer vision, Seismic data processing, Inverse problems, Artificial intelligence, Computer graphics.